(NASA-TM-X-58122) THE ERTS-1 INVESTIGATION (ER-600). VOLUME 6: ERTS-1 SIGNATURE EXTENSION ANALYSIS, JULY 1972 - JUNE 1973 (NASA) 88 p HC $5.00 CSCL 05B
N76-11538
Unclas
G3/43 01570
JSC-08461

NASA TECHNICAL MEMORANDUM

NASA TM X-58122
June 1974

THE ERTS-1 INVESTIGATION (ER-600)

VOLUME VI — ERTS-1 SIGNATURE EXTENSION ANALYSIS

(REPORT FOR PERIOD JULY 1972 - JUNE 1973)

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

LYNDON B. JOHNSON SPACE CENTER

HOUSTON, TEXAS 77058

1. Report No. TM X-58122	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle THE ERTS-1 INVESTIGATION (ER-600): VOLUME VI — ERTS-1 SIGNATURE EXTENSION ANALYSIS (REPORT FOR PERIOD JULY 1972 - JUNE 1973)		5. Report Date 6. Performing Organization Code June 1974
7. Author(s) R. Bryan Erb		8. Performing Organization Report No. JSC-08461
9. Performing Organization Name and Address Lyndon B. Johnson Space Center Houston, Texas 77058		10. Work Unit No. 641-14-07-50-72 11. Contract or Grant No.
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D.C. 20546		13. Type of Report and Period Covered Technical Memorandum 14. Sponsoring Agency Code

15. Supplementary Notes

The JSC Director waived the use of the International System of Units (SI) for this Technical Memorandum because, in his judgment, the use of SI Units would impair the usefulness of the report or result in excessive cost.

16. Abstract

Feature classification, spatially and temporally, was extended over the Houston Test Site Area. The Earth Resources Technology Satellite (ERTS-1) multispectral scanner data from August, September, and October 1972, of five widely separated lakes were used as statistical training fields and test sites.

Short-term temporal (same day to 36 days) and moderately long-term spatial (within and between three ERTS multispectral scanner frames) signature extensions have been verified with respect to large relatively homogeneous features. The most significant feature-dependent variable affecting spatial and short-term extension was water turbidity. Long-term signature extension will require a model to compensate or modify the ERTS-1 multispectral scanner data for significant Sun-angle changes.

The presence of atmospheric haze changed the absolute signature but always by approximately the same amount so that the measured water signature was always the same. The normally occurring variations in atmospheric haze conditions had no major effect on the water signatures in this study.

17. Key Words (Suggested by Author(s)) · Temporal · Atmospheric	18. Distribution Statement STAR Subject Category: ~~18~~ 13

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 88	22. Price

PREFACE

This report is one of seven separate reports prepared by six discipline-oriented analysis teams of the Earth Observations Division at the Lyndon B. Johnson Space Center, Houston, Texas.

The seven reports were prepared originally for Goddard Space Flight Center in compliance with requirements for the Earth Resources Technology Satellite (ERTS-1) Investigation (ER-600). The project was approved and funded by NASA Headquarters in July 1972.

This report (Volume VI) was accomplished by the Signature Extension Analysis Team. The following is a list of the team members.

A. C. Anderson, Lockheed Electronics Company, Inc.
C. R. Hallum, Lockheed Electronics Company, Inc.
C. A. Helmke, Lockheed Electronics Company, Inc.
W. A. Holley, Lockheed Electronics Company, Inc.
A. A. Holth, Lockheed Electronics Company, Inc.
C. J. Liszcz, Lockheed Electronics Company, Inc.
F. W. Solomon, Lockheed Electronics Company, Inc.

The total investigation is documented in the following reports.

Volume	Title	NASA Number
	A COMPENDIUM OF ANALYSIS RESULTS OF THE UTILITY OF ERTS-1 DATA FOR LAND RESOURCES MANAGEMENT	SP-347 JSC-08455

Volume	Title	NASA Number
I	ERTS-1 AGRICULTURAL ANALYSIS	TM X-58117 JSC-08456
II	ERTS-1 COASTAL/ESTUARINE ANALYSIS	TM X-58118 JSC-08457
III	ERTS-1 FOREST ANALYSIS	TM X-58119 JSC-08458
IV	ERTS-1 RANGE ANALYSIS	TM X-58120 JSC-08459
V	ERTS-1 URBAN LAND USE ANALYSIS	TM X-58123 JSC-08460
VI	ERTS-1 SIGNATURE EXTENSION ANALYSIS	TM X-58122 JSC-08461
VII	ERTS-1 LAND-USE ANALYSIS OF THE HOUSTON AREA TEST SITE	TM X-58124 JSC-08463

CONTENTS

TABLES

FIGURES

THE ERTS-1 INVESTIGATION (ER-600)

VOLUME VI - ERTS-1 SIGNATURE EXTENSION ANALYSIS

(REPORT FOR PERIOD JULY 1972 - JUNE 1973)

By R. Bryan Erb
Lyndon B. Johnson Space Center

1.0 SUMMARY

The purpose of the Signature Extension Team was to investigate and assess the feasibility of extending feature classification spatially and temporally over the Houston Area Test Site (HATS) using a minimum number of ground-truth and training field sites. Atmospheric haze and solar elevation angle are the two variables which have the greatest effect on the ability to extend signatures, apart from the variabilities in the targets themselves. The plan adopted by the Signature Extension Team was to collect a library or data bank of signatures, verify them through ground truth, and utilize them for classifying ERTS scenes from other times and other locations. The substantial change in solar elevation angle from season to season forced the data bank to be a function of a calendar date as well as target; that is, the signature depends not only on what is being seen but also on when it is being seen.

Water was selected as the test feature because of its homogeneity over large areas and its invariability over long periods of time. The purpose was to have an easily identified, constant target, so that changes in the signature

could be ascribed to changes in the atmospheric haze and the solar elevation angle. Five water bodies were selected for ground-truth data acquisition, statistical training fields, and test sites. They were Sheldon Reservoir and Lakes Somerville, Livingston, Steinhagen, and Houston. They are widely separated within and near the HATS area to satisfy the need to test spatial extension of spectral signatures. Lake Steinhagen occurs in the overlap between ERTS scenes from 2 consecutive days and provides the data needed for short-term temporal extension using a single target.

The basis or standard data set for this effort was the ERTS-1 multispectral scanner (MSS) data for August 29, 1972, of Lakes Livingston and Houston and Sheldon Reservoir. Extension data sets included the above set plus August 28 and 29, September 15 and 16, and October 3 and 4 for Steinhagen Lake; September 16 and October 4 for Lakes Livingston and Houston and Sheldon Reservoir; and August 30, September 17, and October 5 for Lake Somerville.

1.1 SPATIAL AND SHORT-TERM TEMPORAL EXTENSION OF SPECTRAL SIGNATURES

Water turbidity was determined to be the most significant feature-dependent variable. This parameter varied from 2 to 5 parts per million (ppm) of suspended particles in Lakes Livingston and Somerville to 90 ppm in Lakes Houston and Steinhagen. Applying a semiparametric, untrained, discriminant technique (ISOCLS) to the ERTS-1 MSS data resulted in the generation of seven classes of water, which described the deep areas (over 2 feet) of Lakes Livingston and Houston and Sheldon Reservoir. Seven

additional types of water were obtained from these three lakes due to shallow water, vegetation in the water, and the ratio of water to land in the picture element.

A maximum likelihood technique (LARSAA - Laboratory for Application of Remote Sensing AA) was used to extend these signatures within and between all five lakes for the same day and up to 90 days later. LARSAA is a parametric, trained classification method that utilizes the statistical means, variances, and covariances that describe each class. With one exception, this type of extension (ranging from same-day coverage by ERTS-1 to 36 days later) determined that variations of atmospheric haze were insignificant in water classification, especially when compared to changes in water due to rain and wind direction. The exception was a relatively thick cirrus cloud that covered the western portions of Lake Somerville on August 30, 1972, which increased the apparent brightness of that portion of the lake by a factor of 6.

Another unverified possible exception occurred on the August 28th and 29th coverages of Steinhagen Lake. The apparent brightness of the lake was discovered to have increased during a 1-day period. At that time, no ground-truth effort was being applied to this site. Therefore, it is uncertain whether the change was due to atmospheric haze or some physical change in the lake condition, such as increased wind. The phenomenon was not seen again after a ground-truth effort was established at the site.

1.2 LONG-TERM TEMPORAL EXTENSION

Long-term temporal signature extension using constant signatures was found to be significantly degraded by the change of sun angle. The lower sun angle of late fall and winter caused the data levels of the five sites to drop by as much as 10 MSS units, even in channel 1 (band 4), where random changes are usually one to two units. Thus far, no attempt has been made to compensate for this type of change.

1.3 CONCLUSIONS

A capability to do short-term temporal (same day to 36 days) and moderately long-distance spatial extension of spectral signatures within and between the three ERTS-1 MSS scenes with respect to large, relatively homogeneous features, such as water, has been verified.

Long-term temporal signature extension for the above-mentioned features would require a model to compensate or modify the ERTS-1 MSS data for significant changes of sun angle. Therefore, at present, a data bank approach to signature extension/classification would have to be developed on a seasonal basis.

Normally occurring variations in atmospheric haze conditions appear to have no significant effect on the signatures of the above-mentioned type of features.

2.0 INTRODUCTION

2.1 SIGNATURE EXTENSION

Automatic data classification requires signature extension in some form when the area to be classified was not used to train the classifier. When the training fields are distributed throughout the area to be classified, the signature extension is over a very short distance and not over time. The extension can be thought of as being analogous to an interpolation. However, the question arises as to whether the signatures derived from one area will extend to other areas for which no training data are available. The extension can be spatial where the data are acquired at the same time, but ground-truth data are available for only one area. The extension can be temporal where data are acquired over one area at two different times and the spectral signatures derived from one set of data are used to classify the other set of data. The extensions from one area to another and from one time to another can be thought of as being analogous to an extrapolation.

The most common type of signature extension, as well as the most difficult, involves the simultaneous spatial and temporal extension of spectral signatures. In such an operation, the signatures derived from training data in one ERTS scene are used to classify the data in another ERTS scene of different location acquired at a different time.

2.2 OBJECTIVES

The objectives of the signature extension investigation were to

a. Study the effects of instrument, target, atmosphere, sun elevation angle, and processing variations on the ability to extend feature classification.

b. Evaluate the feasibility of extending feature classification both spatially and temporally over the Houston Area Test Site (HATS) using a minimal number of training sites.

c. Determine what procedures would be necessary to perform feature classification in areas where *in-situ* ground-truth data were not available.

2.3 SCOPE

To achieve efficient utilization of the available resources, the scope of the investigation was limited in several ways. The study was limited to 1 year to produce usable results in a timely manner. The study area was limited to the Houston area to keep the time and expense of gathering ground truth reasonable. The study was limited to three ERTS-1 data sets to allow some depth of analysis on each, rather than a cursory analysis of many data sets. The tools for data analysis were limited to existing computer programs so that the emphasis would be on the ERTS-1 data and not on software generation. The targets

for signature extension studies were limited to fresh water lakes, because they were expected to yield the most information on the variables which could cause identical targets to have different signatures.

2.4 APPROACH

The investigation was conducted by simply extending signatures from one set of data to another and evaluating the results. If an attempt to extend signatures failed, the analysis of the reason for the failure should yield the variables which would prevent the use of universal constant signatures.

The signature spatial extension approach was that of a small step at a time. The initial extension was to be within a given body of water. The next step was to extend to another site within the same ERTS-1 strip. Extension would then be attempted between various test sites in different strips of the same scene. Temporal extension would then be attempted, first by extension to a test site on a preceding or succeeding day, and then by extension over a 36-day time separation to the same site. Additional temporal extensions were to be attempted if the data became available.

3.0 COMPUTER SOFTWARE FOR ANALYSIS

3.1 INTRODUCTION

Because signature extension is a part of automatic computer classification, the investigation was strongly computer oriented. The computer programs which were used in the investigation are described to provide a background for the way they were used. A broad view of the programs is necessary in order that the details of the investigation may be understood in their proper sequence. The previously available programs which were used in this analysis are described in this section.

3.2 SOFTWARE AVAILABLE BEFORE STUDY

Several programs were available before the start of the ERTS project. They were developed primarily to handle aircraft scanner data and they solved a somewhat different set of problems than those presented by the ERTS data. Aircraft scanners used by JSC aircraft have up to 24 channels, and one of the necessary operations in data processing is to reduce the number of channels so that the computation load is reduced.

Small changes were made in all of the programs to adapt them for signature extension use. Generally, they consisted of recompiling a few FORTRAN statements to print out more decimal places, rewind a tape more efficiently, bypass an unnecessary program termination, alter branching criteria, and similar minor changes. No major reprogramming efforts were undertaken.

3.2.1 LARSAA Program

The LARSAA program has four processors which perform different operations on the ERTS data. STAT computes the means and covariances for the class represented by each rectangular training field specified by the investigator. SELECT finds a subset of the data channels which does the "best" job of separating the different classes in spectral space. This reduces the dimensionality of the problem when there are many channels of data, and would hardly ever be used for ERTS data. The number of channels usually is reduced to four. Since the ERTS data contain four channels to start with, there is no great need to reduce the dimensionality of the problem. CLASSIFY uses the information generated by STAT to assign each pixel to a class that is represented by one of the training fields. The output of CLASSIFY is a map tape with each picture element assigned to a class, along with a distance in spectral space from the point to the class mean. DISPLAY uses the map tape as input and prints a map using different characters for the different classes. The investigator may specify a threshold, so that points which are too far from their class mean will be rejected and printed as a blank.

The CLASSIFY processor can be accessed directly by manufacturing a deck of mean and covariance cards to simulate the STAT processor output. This feature permits the use of statistics from one ERTS frame to classify the data in another frame. The process represents direct extension of the signature from one area to another.

3.2.2 ISOCLS Program

The ISOCLS program groups the ERTS data into clusters in spectral space. The clusters have the characteristic that all of the data points included in a given cluster are close together in spectral space. The closeness is obtained by assigning each picture element to the cluster whose center is closest by a simple distance measure. The "volume" of spectral space occupied by a cluster is limited by requiring that the standard deviations in all spectral directions in a cluster be no larger than some limit set by the investigator.

The output of ISOCLS is a map of the area under consideration using different printer symbols for the different clusters. The program will also punch a statistics deck which can be input to the LARSAA-CLASSIFY processor. Signature extension may thus be accomplished by using the ISOCLS output deck from one ERTS scene as the CLASSIFY input deck for another scene. Signature extension may also be performed using only the ISOCLS program by specifying the centers of the clusters for the first iteration from the output of another data set. The format of the card output is not compatible with the card input, but repunching the numbers is not difficult because there are only the four ERTS channels in each cluster, and usually there are no more than 20 clusters.

3.2.3 PICMON Program

The PICMON program produces maps of the ERTS-1 data in individual ERTS channels, and is useful for editing and for

various investigative and diagnostic purposes. Program PICMON can either place the data in histogram form into bins of approximately equal activity to compress the data scale before printing the map, or the investigator may specify the bin edges to suit his purposes. For investigating water, the signature extension team used bins which contained only one of the integer data values per bin to determine exactly where each data level occurred in the map.

3.2.4 REFORM Program

The REFORM program was necessary to reformat the ERTS data into the LARSYS II format accepted by the available processors. Programs LARSAA, ISOCLS, and PICMON all accept the LARSYS II format and not the ERTS format. The version which was first made available to the team was very inefficient and required 45 minutes to convert a complete tape. Thus, the first few conversions that were run were only of selected areas of an ERTS tape. The LARSYS II format includes a line number so that the data could be correlated with the original ERTS data records, which do not contain line numbers. The information in the ERTS header record was lost during the conversion process.

3.3 SOFTWARE DEVELOPED DURING STUDY

3.3.1 Modified ISOCLS Program

The use of ISOCLS to study water signature details was unsuccessful in the early attempts because the data values for water are low (the water appears dark). Since the data values were small, the differences were also

small. If a small allowable standard deviation was entered, the targets other than water were forced to form a large number of clusters, and the program would quickly reach its storage limits and start operating in a degenerate mode. The problem was overcome by allowing the limiting standard deviation to be a function of the data value. Large data values could have a larger standard deviation and small data values were forced to have a small standard deviation. The modified ISOCLS program has been used extensively to study details of the water signature and has found as many as 14 different signatures in a body of water such as Lake Houston, without getting down to the point that each cluster was a quartet of integers with a standard deviation of zero. If the allowable standard deviation is forced to be too small, the clustering routine could degenerate into the individual lattice point mode.

3.3.2 NIAGRA Program

The attempts to study the fine structure of the water signatures revealed residual errors in the data calibration. The mean value returned by each individual detector in each spectral channel was different over a large, homogeneous lake and the clustered output had different clusters arranged in horizontal stripes. Figure 3-1 shows an example of the stripes in the data for Lake Livingston. To provide good visual contrast for the two classes of water which appear within the main body of the lake, one is shaded yellow and the other is uncolored.

NASA S-73-18547

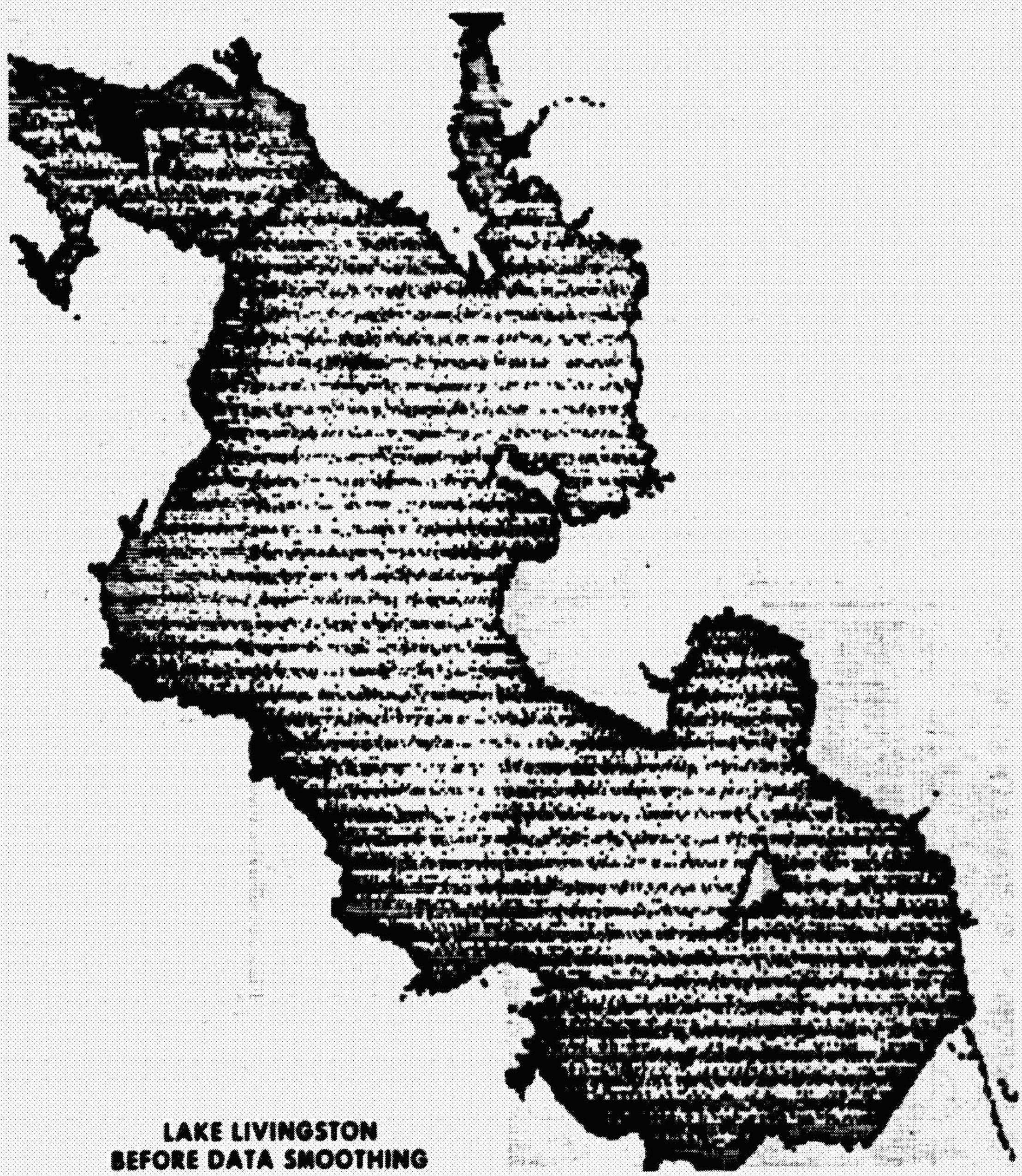


Figure 3-1.- Lake Livingston before data smoothing.

Program NIAGRA smoothes the data and allows the subtle differences in the signatures to be examined without the distracting stripes. The program moves the mean value for each detector to the mean value for all detectors in its channel by changing an occasional data value. The overall average for each channel remains unchanged so that the radiometry is not altered. Because the average value for all detectors in each channel is the same, there are no stripes. Figure 3-1 shows an example of the data for Lake Livingston before data smoothing. The cluster means are exactly the same as for figure 3-2, which shows the lake after data smoothing.

3.3.3 PICTOO Program

Program PICTOO generates a two-dimensional histogram of data from two ERTS-1 channels. The histogram, which is in the form of a table, gives a picture of overall ERTS data structure in two-dimensional spectral space. Gray-level maps and cluster maps indicate local ERTS data structure, which may be related to ground-truth information. However, data levels and cluster signatures vary between similar ground features within the same ERTS pass, and between the same ground feature on different passes. A two-dimensional histogram may be used to look for characteristics of the data, which are relatively invariant and, therefore, of possible use in signature extension.

The table generated by PICTOO contains an entry for each combination of data values from two ERTS channels. For example, if channels 1 and 4 are under study, then the *ith*,*jth* entry in the table gives the number of pixels for which channel 1 has a data value of i and channel 4 has a

NASA S-73-18546

Figure 3-2.- Lake Livingston after data smoothing.

data value of j. The values i and j range from 0 to 127 in the case of channels 1, 2, and 3, from 0 to 63 in the case of channel 4.

Program PICTOO will generate a two-dimensional histogram for a set of rectangular areas. Each area is defined by the beginning and ending line and pixel numbers.

3.3.4 EDICFF Program

Program EDICFF provides a very general method of selecting ERTS-1 data for statistical analysis. The data selected are written on tape in a convenient format for input to a statistical analysis program, such as the UCLA BMD statistical package. For these applications, determining useful data sets from gray-level maps and cluster maps is often difficult because they are defined by a collection of rectangular areas. If this is attempted, the areas may turn out to be quite small, and in some classes, degenerate. The method of selection employed by EDICFF is to specify a starting point in the data (a line number and sample number), a count of picture elements to be selected beginning at that point, and a class symbol to be associated with the data.

From this information the EDICFF program creates a symbol or cluster map representation of the data to be extracted from the ERTS-1 data tape. The program then reads the ERTS-1 data tape line by line and consults the symbol map to determine the picture elements for which data values are to be saved. After data selection has been completed, the program sorts the data by class symbol

in an order specified by the user, and writes the data on tape. The program also provides a listing of data values by class and a printout of the class symbol map.

3.3.5 TMERGE Program

The TMERGE program combines or merges two ERTS-1 data tapes which contain data from two adjacent strips of an ERTS frame. The tapes are in the LARSYS II format. The purpose of merging tapes is to study, for example, by clustering, ground features such as lakes, which are in two strips. If the two input tapes have 810 samples per scan line, the output tape will contain twice that number, or 1620 samples per line. The merge is accomplished without unpacking the data and is, therefore, not very time consuming.

3.3.6 Fast PICMON Program

Program PICMON, which is used to generate a gray-level map of the data from an ERTS-1 channel, was modified to reduce program running time. The change enables the program to run in less than a third of the time originally required. This savings in time is important, since the program is frequently used as the first processing step in examining a large amount of ERTS data. The improvement was obtained by a change in the method of unpacking the ERTS input data. ERTS data are stored on tape, one record per scan line, in which the many 8-bit data values making up a record are stored in a packed format. The original method of unpacking the data was by means of a separate subroutine call for each 8-bit data value. This inefficient procedure was replaced by a single subroutine call to

unpack the data for a line, and then to unpack only the data for the sample interval and the channel for which the gray map was being generated.

3.3.7 Fast REFORM Program

The original REFORM program required too long to run and some changes were made to speed it up. The major change was in the unpacking and repacking of the data to get from one format to the other. The more efficient handling of the two operations decreased the running time for one 25- by 100-n. mi. strip from 45 minutes to 15 minutes on the Univac 1108. The identification and header record information were still lost, but the data records were available and no further improvements were attempted. The correct method of solving the problem of different formats for any production work is to rewrite the read and unpack routines for the processors so that they can take the ERTS tape directly.

3.3.8 FEOW Program

Based on the experience gained by studying and evaluating the signatures of various types of water and mixture picture elements, a classification program was written called FEOW. At present its capabilities are limited with regard to the sun angle or season of the ERTS pass from late spring through early fall. It has only been used to evaluate fresh water, and the data must be in the form produced by using program EDICFF.

Program FEOW first determines if there is any fresh water in the picture element by checking the channel 4 data

value. If not, it stores a blank to be printed out for that space. If there is water in the picture element, it next determines if it is a mixture or a total water sample by subdividing the acceptable channel 4 data range. If it is totally water, it calculates the level of turbidity, using the channel 1 data value, within the range of 2 ppm to 100 ppm of suspended solids as shown in table 3-I. If it is an edge picture element, it calculates the approximate percentage of water in the sample by further subdivision of the channel 4 data range as shown in table 3-II.

Finally, FEOW generates a gray map of the scene. Figure 3-3 is an example of the output for Lake Houston on August 29, 1972. Figure 3-4 is the output for Steinhagen Lake on August 29, 1972.

TABLE 3-I.- SYMBOLS AND COLOR CODE FOR FIGURES 3-3 and 3-4

TOTAL WATER PIXELS			EDGE PIXELS	
SYMBOL	COLOR	TURBIDITY	SYMBOL	%WATER
0		0 TO 9 PPM	A	1-10
1		10 TO 19 PPM	B	11-20
2		20 TO 29 PPM	C	21-30
3		30 TO 39 PPM	D	31-40
4		40 TO 49 PPM	E	41-50
5		50 TO 59 PPM	F	51-60
6		60 TO 69 PPM	G	61-70
7		70 TO 79 PPM	H	71-80
8		80 TO 89 PPM	K	81-90
9		90 TO 99 PPM	J	91-99
		BORDER BETWEEN TOTAL WATER AND EDGE PIXELS		

TABLE 3-II.- MIXTURE PICTURE ELEMENTS

Symbol	Percent Water	Channel 4 data value
A	1-10	15
B	11-20	14
C	21-30	13
D	31-40	12
E	41-50	11
F	51-60	10
G	61-70	9
H	71-80	8
J	81-90	7
K	91-99	6

Figure 3-3.- Lake Houston turbidity map for August 29, 1972.

Figure 3-4.- Steinhagen Lake turbidity map for August 29, 1972.

4.0 SIGNATURE EXTENSION INVESTIGATION

4.1 INTRODUCTION

The investigation of signature extension involves the study of all sources of variability in the data. These include both instrument and data processing factors and target-related factors. This section discusses the data characteristics which are related to the instrument and the data processing, followed by a discussion of the experiments which revealed target-caused variabilities in the ERTS data.

4.2 STUDY SITES

The several bodies of water in and around the HATS area selected as study sites were Lakes Somerville, Livingston, and Houston, Sheldon Reservoir, and B. A. Steinhagen Lake.

The sites were selected for their size, location, and varying physical characteristics. The physical characteristics of the sites which most affected the spectral signature were turbidity (suspended particles), depth, and vegetative growth (both floating and rooted). Certain sites were relatively homogeneous over their main body, while others varied greatly over their entire length. A brief description of each of the study sites follows.

4.2.1 Lake Somerville

Lake Somerville is a Corps of Engineers impoundment on Yegua Creek, a tributary of the Brazos River (latitude and longitude 30°19' N., 96°35' W.). The approximate size of the main body of the lake is 14.4 kilometers by 3.2 kilometers (9 miles by 2 miles). The lake bed was cleared of trees prior to the filling of the lake, with the exception of certain shallow areas and the upper end of the lake. Yegua Creek is a minor tributary and does not have great length, which results in a low turbidity level in the lake. The lake is very homogeneous over its length, and signature variation occurred mainly in the upper end where standing trees extend above the water's surface, and in the coves where surface vegetation occurs. This site was chosen initially because it would appear in the overlap of 2 consecutive days' coverage. The satellite orbit was shifted after insertion and the expected overlap did not occur.

4.2.2 Lake Livingston

Lake Livingston is a Trinity River Authority impoundment on the Trinity River, a major watershed extending from north of the Dallas/Fort Worth area to the Gulf Coast. The approximate size of the main body of the lake is 25.6 by 8 kilometers (16 by 5 miles) as shown in figure 4-1. This is the largest and most homogeneous of the study sites (latitude and longitude 30° 43' N., 95°08' W.). The lake bed was well cleared of trees prior to the filling of the lake, except for the upper portion above the U.S. 190 causeway and bridge. In this area there were many standing trees

NASA S-73-25514

LAKE LIVINGSTON

ERTS-1 TEST SITE

UNCONTROLLED MOSAIC
SCALE 1:120,000

Figure 4-1.- Lake Livingston ERTS-1 test site.

whose crowns extended above the surface of the lake. The lake has a low turbidity level and affords the opportunity for signature extension to Lake Somerville. The size of this target also permits extension experiments within a test site. The geometric properties of the location of this lake with respect to the ground track cause the lake to appear on two adjacent strips of an ERTS-1 scene.

4.2.3 Lake Houston

Lake Houston (figure 4-2) is a City of Houston project and is an impoundment of the San Jacinto River. The approximate size of this lake is 14.4 by 3.2 kilometers (9 by 2 miles). This lake is of great interest because of the varying turbidity levels in different areas of the lake (latitude and longitude 29°60' N., 95°08' W.). At the upper end of the lake are the east fork and west fork of the San Jacinto River. The east fork has a moderate turbidity level and the west fork has a high turbidity level. The two streams enter into a "mixing bowl" area with a turbidity level between those levels found in the individual forks. This mixing bowl area occurs to the south of the McKay Bridge and causeway (Atascosita Road). The turbidity level decreases as the water flows down the lake and the suspended particles settle out. The mixing area increases in size as the flow rate of the west fork increases.

NASA S-73-25526

Figure 4-2.- Lake Houston ERTS-1 test site.

4.2.4 Sheldon Reservoir

Sheldon Reservoir is a Texas Parks and Wildlife impoundment whose main purpose is to provide a wintering location for migratory waterfowl. It is a shallow reservoir with a high level of aquatic vegetation and a moderate level of turbidity (latitude and longitude 29°52' N., 95°11' W.). It is an impoundment of Carpenters Bayou, a tributary of the San Jacinto River. Sheldon Reservoir is located 5 kilometers (3 miles) south-southwest of Lake Houston and affords a convenient site for short-distance signature extension experiments. Its spectral signature, however, differs significantly from that of the major portion of Lake Houston.

4.2.5 B. A. Steinhagen Lake (Dam B)

The B. A. Steinhagen Lake is a Corps of Engineers impoundment on the Neches River (latitude and longitude 30° 53' N., 94° 11' W.), with inflow also from the Angelina River. The approximate size of this impoundment is 9.7 by 3.2 kilometers (6 by 2 miles). This lake was not initially a study area, but was chosen later because it appeared in the overlap on 2 consecutive days of coverage by ERTS-1 after its adjusted orbit. It replaced Lake Somerville in importance in the consecutive days' extension of a study site. (This experiment could not be attempted because of the unacceptable cloud conditions during ERTS-1 overpasses.)

The lake, similar to Lake Houston, is an interesting study of variations in turbidity over a body of water.

Relatively clear water enters the impoundment, which is shallow with a muddy bed, and the turbidity levels increase as the water flows down the lake. As wind conditions increase, the turbidity levels of the lake increase as additional particles are placed in suspension. The lake, similar to Lake Houston, affords an opportunity of extending signatures between the two sites.

4.3 ERTS-1 DATA CHARACTERISTICS

Some characteristics of the ERTS-1 data have detrimental effects on the statistical processes used in the available computer programs. The first of these is that the data values are integers. The discrete rather than continuous nature of the data has an effect on the value of the standard deviation of a cluster of data points, especially when the standard deviation is of the same order of magnitude as the separation of the discrete values. There are, however, some anomalies in the data that are even more destructive to the meaning and interpretation of statistical results. These include geometric distortions of the grid of data values, preferred and missing data values, and incomplete calibration of the data.

4.3.1 Geometric Distortions

Line-printer maps produced by programs PICMON, ISOCLS, and LARSAA contain geometric distortions which make correlation with airborne photography and standard maps very difficult. The ERTS-1 data contain a differential scale which is different from the differential scale of the line printer. Each data vector represents a rectangle

on the ground whose edges are on the ratio of 56:79. The line printer reproduces the scene on a grid of rectangles whose edges are on the ratio of 3:5. Since the two ratios are not equal, the line-printer map is stretched in one direction with respect to the other.

The rotation of the earth during the time required to complete a scan causes each scan to be offset from the next. Thus, a skew is introduced into the grid, which appears in the line-printer map.

4.3.2 Preferred and Missing Levels

When the ERTS-1 data were examined on a microscopic level, certain data values were discovered to occur far more frequently than others. The data from a homogeneous target would be expected to be distributed about a mean value with a Gaussian distribution. In fact, certain values appeared much more often than they should. If the data are examined on a detector by detector basis, the anomalies are even more drastic. Every sixth line through the frame was measured by one specific detector; and if every sixth line was taken as the data set, there were data values which occurred much too frequently, while the next higher or lower data value did not occur at all. Such systematic unevenness in the distribution of the data values tends to subvert any meaning which might be attached to variances and standard deviations. This unevenness also moves the mean value for a homogeneous feature away from the center of a distribution as determined by the shape of the wings of the distribution.

4.3.3 Incomplete Calibration

An examination of the data for the large lakes studied by the signature extension team revealed that certain detectors gave consistently high readings, while others gave consistently low readings. The differences are attributable to a residual error in the calibration. The error can take two forms, either offset or gain. The offset error is independent of the data level and appears as a constant added to or subtracted from every reading from a given detector. The gain error appears as a wrong slope for the data value versus scene radiance line. Both types of error are present in the data, and they tend to increase the standard deviation for the data belonging to a given class in the scene.

The addition of a miscalibration component to the standard deviations further subverts any physical significance that might be attached to them. As stated earlier, all of the available processors use the standard deviations as the unit of measure in spectral space.

4.4 SIGNATURE BEHAVIOR

A complete understanding of the way the signatures behave is a prerequisite for performing signature extension on a routine basis. All of the factors which can cause a signature to change must be identified and their influence must be considered when performing identification by signature extension. By using water as a target, the influence of changes in the target itself was minimized, and the influence of the scanning instrument, the data

processing, the atmosphere, and the solar elevation angle could be studied. Although the target variability was minimized, it was not eliminated. Therefore, it was also necessary to study the variability within the water targets themselves.

4.4.1 Variables Affecting Signature

The following variables were initially considered in planning the study of variability of the signatures of water targets:

a. Temperature
b. pH Factor
c. Turbidity
d. Suspended Solids
e. Atmospheric Haze
f. Sun Angle
g. Bottom Features
h. Standing Vegetation
i. Surface Vegetation
j. Pollution
k. Wind (Surface Condition)
l. Color
m. Chlorophyll A
n. Algae
o. Land (Island or Shoreline)
p. Floating Materials
q. Depth

The measurements of the first six of these variables were taken during ground-truth expeditions at the time of the ERTS-1 passes. The results of signature variability studies have indicated that the following characteristics of the water had the greatest impact.

a. Turbidity: This variable has been the most critical in extending generalized signatures. In two of the study areas (Lake Houston and B. A. Steinhagen Lake) the main body of the lake had several signatures, depending on the turbidity level of the specific area. These areas of turbidity change from time to time as a result of wind, rain, and lake level.

b. Depth: The signatures of most of the study areas change as the scanner passes over the upper reaches of the impoundments or the extremities of the coves.

c. Standing Vegetation: In some impoundments in this geographical area, the trees had not been cleared prior to the filling of the reservoir and this has resulted in trees (both live and dead trunks) protruding through the surface of the water. The signatures vary in these situations.

d. Surface Vegetation: A major problem in some of these impoundments was the introduction of the water-hyacinth and other aquatic surface plants. When these appear, they introduce variability in the signature.

e. Land: Great variability occurs when a picture element contains both land and water. This occurs mainly along a shoreline or for a small pond, and the signature level increases as a result of the ratio of land to water.

These variables are not independent; in fact, they may be highly dependent on each other. For example, in the shallow area of a lake there may be variability caused by depth; standing vegetation, since it is able to protrude through the water's surface; turbidity (shallow areas are more prone to sediment being disturbed as weather changes); and surface vegetation, if the shallow area is somewhat protected.

Water temperat·re measurements will become more important when thermal channels are added to future satellite sensor systems.

4.4.2 Atmospheric Corrections

Lake Livingston.- This lake was selected for the first test of evaluating changes in spectral signature due to atmospheric conditions because of its size and relatively constant turbidity level. Numerous readings were obtained around the lake during two ERTS-1 passes 18 days apart.

The results of this test were inconclusive, because no significant change in the water signature was detected between the two sets of data. A further hindrance in the data evaluation was a lack of information concerning the accuracy and precision of the solar photometers. This problem has still not been corrected, although numerous attempts have been made.

Steinhagen Lake (Dam B).- The initial information concerning the track of ERTS-1 indicated that Lake Somerville would appear in the overlap area of the ERTS-1 pass on 2 successive days. The plan was to use this condition to evaluate changes in atmospheric variation by assuming that the water characteristics would not change significantly in a 24-hour period. Hence, any change in the spectral signature of the lake would be due to atmospheric changes. However, the actual track was off from the proposed track by about 50 miles, and Lake Somerville could not be used for this part of the study.

When the first ERTS-1 imagery was received, Steinhagen Lake appeared in the upper northeast corner of the Lake Livingston scene. Further study showed that it fell in the overlap area of 2 successive days. This part of the

study was then shifted from Lake Somerville to Steinhagen Lake.

Plans were made to attempt an atmospheric haze correction if it were possible to obtain data over a single target for which there was coverage for consecutive days, as well as measurable haze differential. Initially, equipment was not available which would permit the measurement of either the optical depth of the atmosphere or the turbidity of the water. When the equipment became available, there were not 2 consecutive days of clear weather while the ground-truth effort was applied to the Steinhagen study site.

The only good 2 consecutive days' coverage for which data were available was from August 28 and 29, 1972, but no supporting data of water or haze conditions were available. In retrospect, there should have been little change in the features of the site, since there had not been any significant climatic condition preceding these passes that would have affected their signature. If the assumption is made that there was no change in the feature, any change in the spectral signature would have been caused by a change in the optical depth over the 2 days.

Since Steinhagen Lake is variable over its various parts, selected areas were chosen to study the change in reflectance over the 2 days. Twelve areas indicated

similar trends in MSS data values with average changes as follows:

Channel	MSS band	Overall difference
1	4	2.6933
2	5	2.8055
3	6	2.6875
4	7	.8854

The least and most turbid areas had the following changes:

	Least turbid			Most turbid		
Band	8/28	8/29	Change	8/28	8/29	Change
4	23.444	26.0139	2.5695	34.5000	37.000	2.5000
5	14.4583	17.9306	3.4723	27.8333	30.8472	3.0139
6	8.9722	11.3333	2.3611	13.1667	16.6111	3.4444
7	1.5000	2.2500	.7500	2.2083	3.3056	1.0973

Although there was no attempt to adjust the data levels to either each other or to a nonatmospheric basis, it would seem that the various areas of the lake would extend to the same area on the next day if the manipulation had been performed.

There was an attempt to extend the water signature without correcting for the change in atmosphere with the results shown in figures 4-1 and 4-3. Figure 4-4a is the ISOCLS run of the 28th and figure 4-4b represents the statistics of the 28th applied to the data for the 29th. The most turbid area (blue) is larger because of higher

S-73-29854

(a) (b)

Figure 4-3.- 24-hour backward signature extension on Steinhagen Lake.

S-73-29855

(a) (b)

Figure 4-4.- 24-hour signature extension on Steinhagen Lake.

reflectance. Figure 4-3 is the reverse process, with figure 4-3a being the ISOCLS output of the 29th and figure 4-3b being the statistics of the 29th applied to the data obtained on the 28th. The turbid area has shrunk because reflectance levels have decreased.

Lake Somerville.- Although the consecutive day coverage was transferred from Lake Somerville to Steinhagen Lake, Lake Somerville was retained as a study site for the extension across adjoining ERTS-1 scenes. Although the required equipment had become available to do the job properly (January 1, 1973), the weather was bad for each ERTS-1 pass. Therefore, no ground-truth data were acquired for this lake. However, it was apparent after evaluating the ERTS-1 imagery for other lakes in conjunction with ground-truth data, that this lake was very homogeneous and had a low turbidity. Therefore, on August 30, 1972, when ground truth was being gathered on Lake Somerville, and a high thin cirrus cloud covered the western half of the lake at ERTS-1 pass time, there was a chance to have the type of haze data that were needed. Unfortunately, some of the instruments were not in place at the right time. Of those that were, some were unable to obtain stable readings, and those that were able to, did not yield data that concurred with the final ERTS-1 data (see figure 4-5 for an ISOCLS cluster map of the August 30th scene). If the solar photometer data had been good, most of the western portion of the lake (all of the nonyellow) would be the same as the eastern portion after being corrected by an atmospheric model.

The line dropout which was very evident in the image (figure 4-5) was introduced by the Goddard processing.

NASA S-73-25520

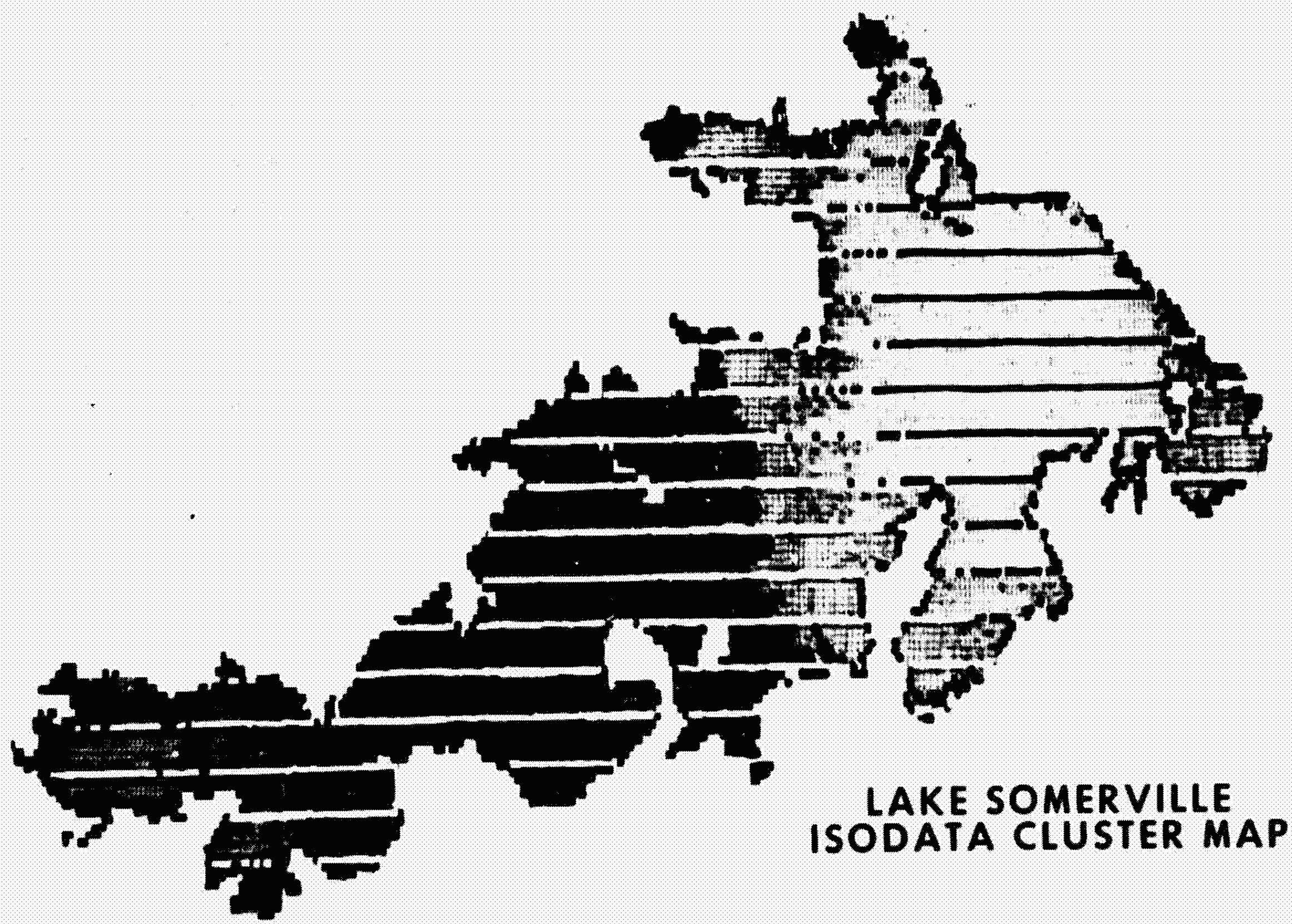


Figure 4-5.- ISOCLS cluster map of Lake Somerville on August 30, 1972.

In a subsequent computer tape from Goddard of the same image, this striping was completely removed.

4.4.3 Seasonal Changes

The presence or absence of certain targets and the appearance of most natural targets depend upon the season. Crops will be present only during the growing season and will not be present during the rest of the year. Forests, grasslands, and brushlands will change their appearance during the year. The only features which will remain relatively constant are water, bare soil, and manmade features, such as large areas of concrete or rooftops. The deep clear lakes remain constant within a data level or two throughout a season. The turbid lakes change in appearance with turbidity, which does change, but not seasonally. Periods of heavy rain will increase the turbidity of the waters, but the heavy rains correlate only approximately with the seasons.

4.4.4 Sun Angle

The sun elevation at 9:30 a.m. ranges between 30° and 60° for the Houston area during the year, which changes the scene illumination by a factor of 1.7 at the time of the ERTS-1 overpass. If the scene were a perfect diffuse reflector, the measured radiance would also change by the same 1.7 factor. However, most features of the scene are not perfect reflectors, and no simple correction is available to normalize to some fixed solar elevation angle. The data level for water is dependent upon the sun angle in the visible channels, but not in the infrared channel. The

targets, such as foliage, which are characterized by multiple reflection are not Lambertian and a cosine correction is not applicable. Bare soil is probably Lambertian and a cosine correction can be applied. Since the sun elevation is perfectly correlated with the calendar date, the correction may be included in the signature for a given date. Indeed, sun elevation probably cannot be separated from the other effects of the seasonal variations in the target.

4.4.5 Correlation of Turbidity With Photometer Data

Ground-truth data were obtained on Lake Houston February 25, 1973, using a Hellige turbidometer to measure water turbidity and five solar photometers to gather atmospheric data, as well as a photometer to measure the target radiance in the ERTS bands without an intervening atmosphere. The weather was good and a fairly high-quality set of ground-truth data was gathered.

Based upon previous data, 17 sample sites were selected at which data on turbidity and corresponding ERTS-1 photometer readings were obtained (figure 4-6). Solar photometer measurements were also made at five locations along the main body of the lake.

Unfortunately, the ERTS-1 MSS data for the same date did not arrive in time to be fully analyzed for this report. The following are the results of the correlation study of the measured values of turbidity and the readings made with the ERTS photometer, and an estimate of correlation with the August 29, 1972, ERTS-1 data. The BMD02R, UCLA

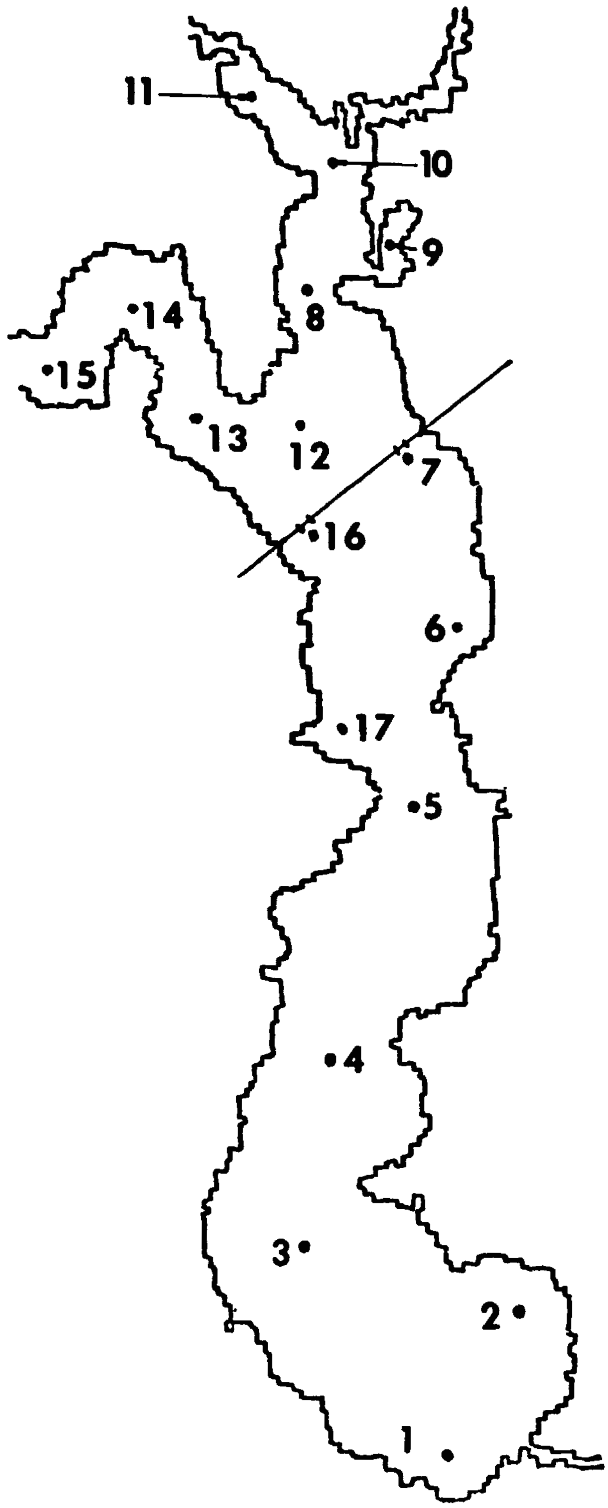


Figure 4-6.- Lake Houston test site locations.

biomedical statistical package program was used, which computes a sequence of multiple linear regression equations in a stepwise manner.

The model was defined at the time of input to the statistical program with the measured turbidity as the dependent variable, and the values recorded for the four channels of the ERTS-1 photometer as the four independent variables.

The first step in the solution of this model indicated that channels 1 and 4 were the most significant of the independent variables. The correlation between the photometer readings and turbidity was 0.95 for channel 1 and 0.96 for channel 4. Using channel 1 or 4 to predict turbidity yielded a standard error about the prediction of ±4.06 ppm (parts per million of suspended solids) over a range of zero to 100 ppm. Using both channel 1 and channel 4 increased the correlation coefficient to 0.97 and decreased the standard error of the estimate to 3.45 ppm. Incorporating the remaining two channels (2 and 3) proved to be statistically insignificant.

4.5 EXTENSION EXPERIMENTS

Several signature extension experiments were performed using various combinations of the programs described in section 3.0.

After tape conversions, the initial step in the investigation of any training site is the production of a density slice (Program PICMON) of the infrared channel for the study site area. Picture elements with gray-level readings in the 0 to 5 range represented the major body of water in the test site. If additional information was required, the gray-level range was increased to a level of 10 or more to bring in "edge" picture elements and smaller ponds. Once the location of the site had been verified, either the clustering algorithm ISOCLS or the training field selection technique LARSAA-CLASSIFY was used.

The investigation routine began with ISOCLS to gain information on the number of classes of water and also statistical information (means and covariances) on these classes. These statistics were introduced into the LARSAA-CLASSIFY algorithm as artificial training field statistics. The ISOCLS identification as to the number of classes gave an indication of the number and location of training fields to increase the identification percentage. The results of the signature extension experiments are described in the following sections.

4.5.1 Signature Extension Study Using Lake Houston as a Target

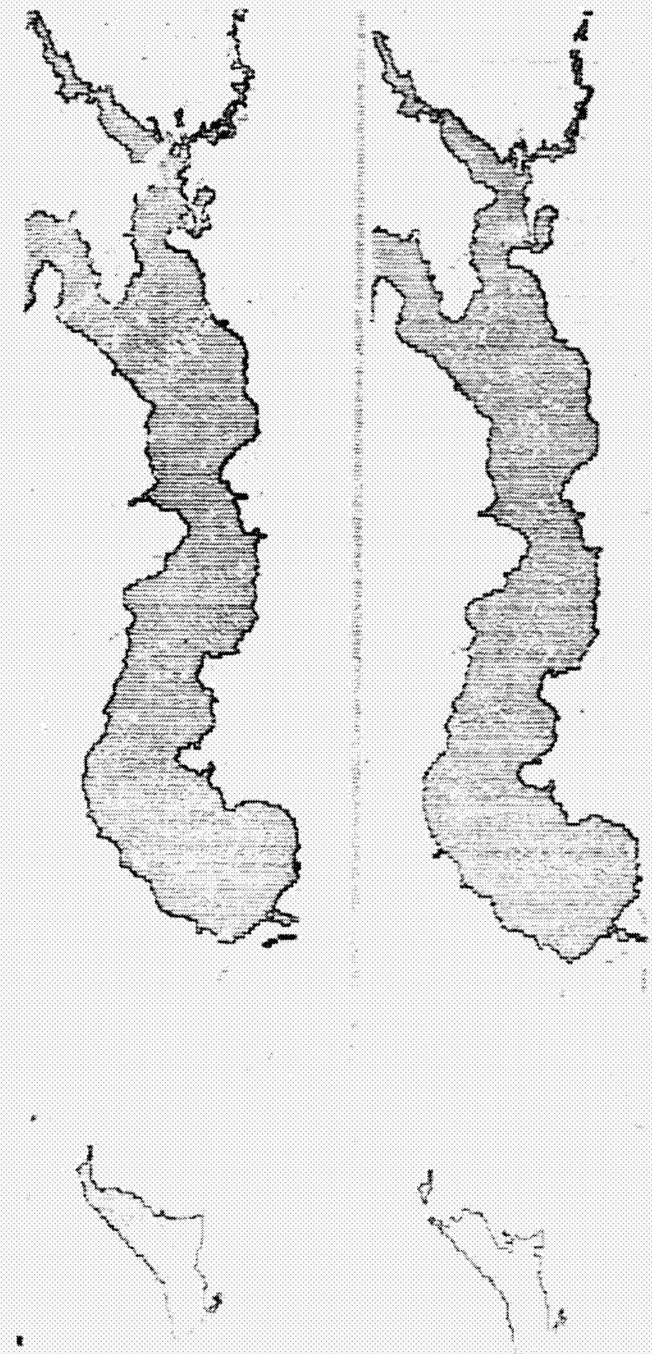
As an illustration of the technique of clustering followed by classifying, Lake Houston and its companion lake, Sheldon Reservoir, were selected as a primary site. As previously indicated, Lake Houston has varying turbidity levels, and Sheldon Reservoir is shallow with much aquatic vegetation.

The initial computer printout is a density-sliced gray map of the lower reflectance levels of the infrared channel using PICMON (figure 4-7a). This provided the location and outline of the lake, which permitted an ISOCLS printout to be obtained of the area. The ISOCLS printout (figure 4-8a) indicated 14 classes of water in the two impoundments, of which five were major classes, three were minor classes, and six classes were mainly "edge type". The number of picture elements and their location were comparable to those produced under the density-slicing technique.

LARSAA-CLASSIFY was then used with artificial training-field statistics taken from the 14 classes of water identified by ISOCLS. Thresholds of 10.0 (5σ) and 2.6 (1σ) were used with the results at a threshold of 10, which represented a 1-percent variance in the number of picture elements identified as water, and a 5-percent shifting of individual picture elements between classes. The results under the 2.6 threshold were a 74-percent identification of water picture elements in the overall scene.

The use of the LARSAA-CLASSIFY was then shifted to the use of actual training fields, and the results achieved were compared with the results of the ISOCLS output. Training-field selection was first attempted by assuming ignorance about the water feature and selecting a training field that would be expected to represent the entire site. A large training field was selected in the main body of the lake and the classification results were not impressive. At a threshold of 10, 71 percent of the water in the scene was

S-73-29859



(a) October 4, 1972 (b) August 29, 1972

Figure 4-7.— Density-sliced gray map of Lake Houston.

S-73-29863

(a) Aug. 29, 1972 ISOCLS

(b) Aug. 29, 1972 LARSAA

(c) Oct. 4, 1972 LARSAA

(d) Oct. 4, 1972 ISOCLS

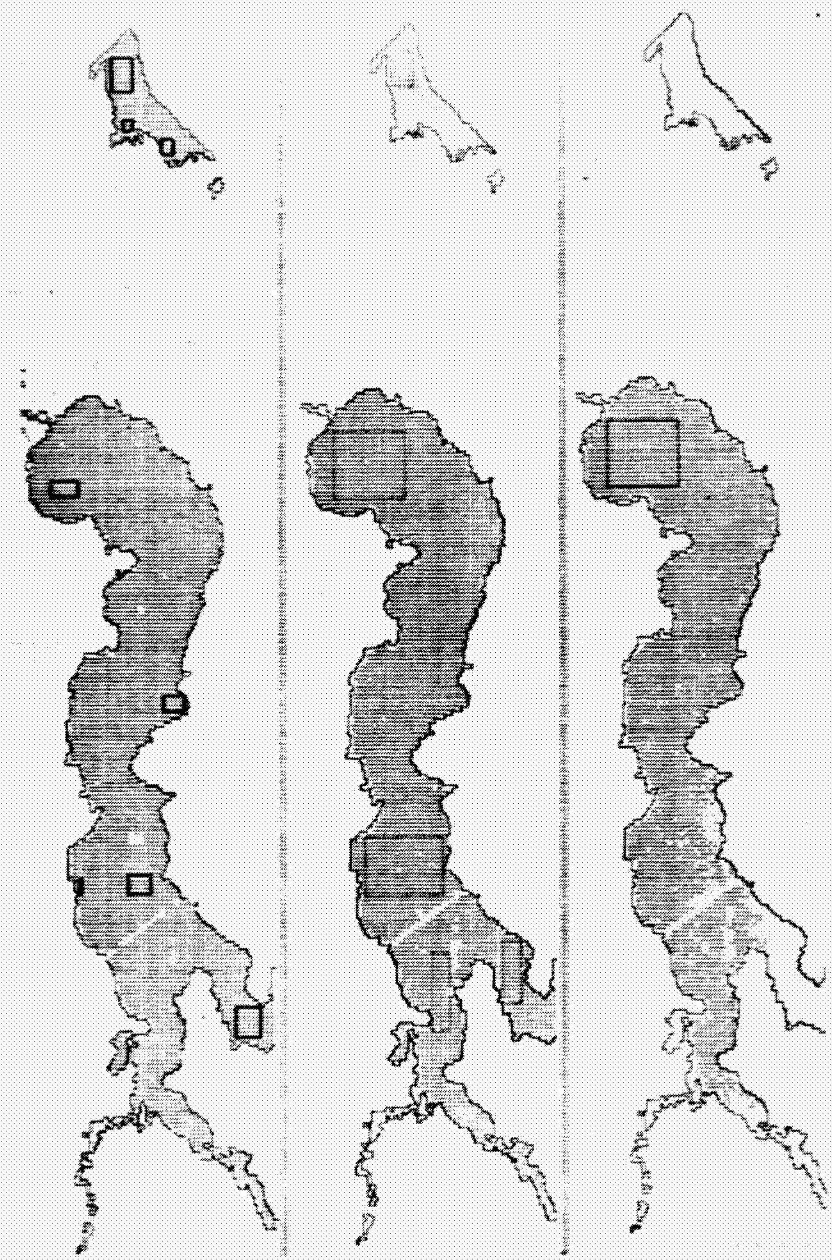
Figure 4-8.- Clustering and classification results for Lake Houston. (Threshold = 10; statistics from ISOCLS of same date.)

identified (shown in figure 4-9a), and at a threshold of 2.3, 37 percent of the water picture elements were identified.

Additional training fields were then selected, one field for each of the five major classes of water as indicated by the ISOCLS output. Two approaches were used: the training fields were used as separate classes of water, and there were assumed to be five training fields for the same type of water. As anticipated, the results were an improvement over the single training field approach. The results using the training fields as examples of a single type of water were 84 percent at a threshold of 10 (shown in figure 4-9b), and 62 percent at a threshold of 2.3. Using the training fields as separate classes of water resulted in a 90 percent identification at a threshold of 10 (shown in figure 4-10a), and 60 percent at a 2.3 threshold.

The areas of the lakes which were not identified were those of the extremes in turbidity level, the turbid west fork and the low-turbidity Sheldon Reservoir. The major tributaries were also relatively poorly identified.

The next extension exercise involved the selection of training fields for eight classes, five major and three minor classes. Again, these were used as separate classes and then combined and used as one class of water. The identification results at a threshold of 10 were 94 percent when used as a single class of water (shown in figure 4-9c), and 95 percent when used as separate classes (shown in figure 4-10c). At a 2.3 threshold, the results were 71 percent and 51 percent, respectively. The majority



(a) Training Field

(b) Training Fields

(c) Training Fields

Figure 4-9.- Classification results for Lake Houston and Sheldon Reservoir using a single class to identify water. (LARSAA, threshold = 10, data of Aug. 29, 1972).

S-73-29862

(a) (b) (c) (d)

Figure 4-10.- Signature extension results for Lake Houston using multiple water classes.

of missing picture elements were of the "edge cell" variety. This group poses a problem for training field selection because of the sporadic nature of their location.

This was the extent of the extension experiment within a given body of water and its neighboring reservoir. The extension experiments then shifted to the extension of signatures to and from the other test sites on the same day (Lake Livingston and Steinhagen Lake), the preceding day (Steinhagen Lake), and the subsequent day (Lake Somerville).

4.5.2 Signature Extension to Other Sites

Extension from Lake Houston was first attempted to Lake Livingston, a distance of 50 miles. The statistics of the eight training fields of Lake Houston were artificially entered as training field statistics under LARSAA-CLASSIFY for the Lake Livingston site. The results at a threshold of 10 were disappointing. Less than 1 percent of the water picture elements in the Livingston scene were identified (figure 4-11). These were the edge type picture elements (most turbid areas), which were identified as comparable to the least turbid areas of Lake Houston.

The extension from Lake Houston to Lake Somerville provided the same results. Edge-type picture elements were partially identified by the statistics of the least turbid training fields. This was an attempt at a 1-day extension over a distance of 100 miles.

Figure 4-11.- Signature extension results for Lake Livingston using statistics for eight classes from Lake Houston.

Extension from Lake Houston to Steinhagen Lake, both collected on the same day, provided better results. These lakes are similar in their levels of turbidity, and extension was anticipated to cause no serious difficulty. The extension was approximately 90 percent successful at a threshold of 10, and 50 percent at a threshold of 2.3 (figure 4-12). The area causing the greatest problem was the area of highest turbidity on Steinhagen Lake.

Extension in the opposite direction was then attempted. The extension results from the other sites to Lake Houston were expected to be similar to the extension from Lake Houston, and this was correct.

The statistics used for extension from Lake Somerville were ISOCLS statistics for the two main classes of water which did not include picture elements with marine vegetation or those described as "edge cells". These two classes were able to identify only 6 percent at a threshold of 10, and 4 percent at 2.3, shown in figure 4-13c. Only the two least turbid areas in the Houston scene (Sheldon and Houston's East Fork) were identified.

The results from the extension from Lake Livingston to Lake Houston were similar, with the exception of starting with five classes from ISOCLS, of which two were main areas, one a shallow area, and the two others were edge picture elements. One of these edge cells caused much misclassification error at a threshold of 10, but no error at 2.3. The two main classes did not identify any picture elements on Lake Houston, and any identification of main sections of

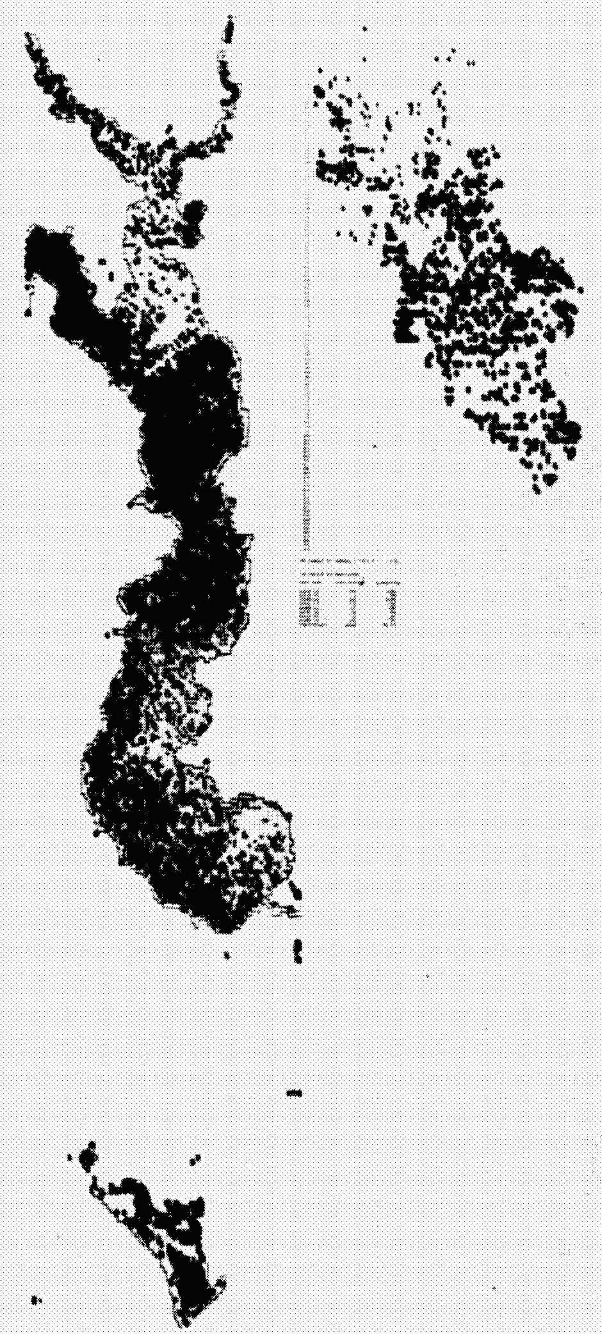
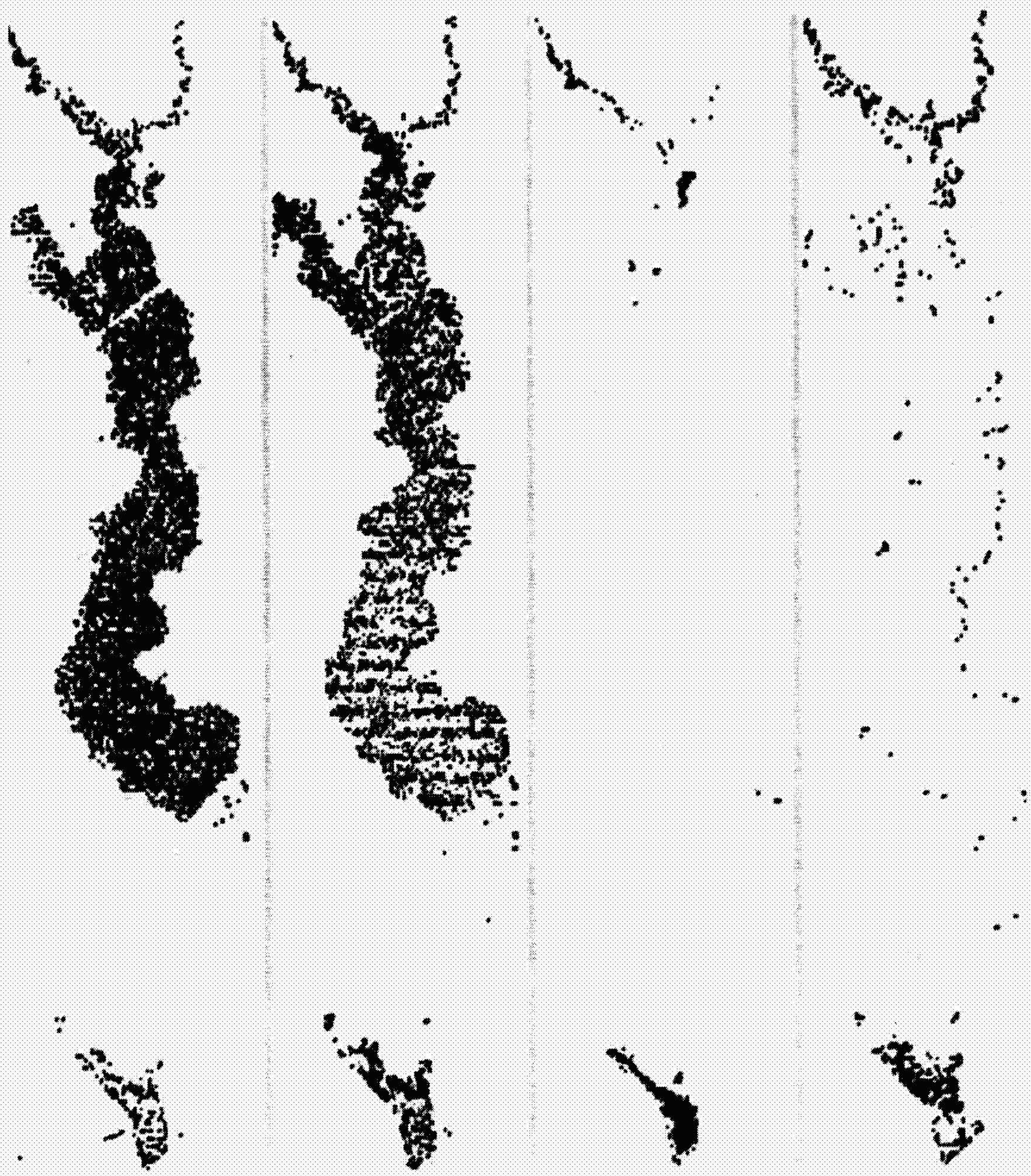


Figure 4-12.— Signature extension results for Lake Houston using Steinhagen Lake statistics and for Steinhagen Lake using Lake Houston statistics.

S-73-29857



(a) (b) (c) (d)

Figure 4-13.- Signature extension results for Lake Houston data of August 29, 1972 (LARSAA, threshold = 2.3).

the lake resulted from the statistics acquired from the edge picture elements from Livingston. The overall results were 58-percent identification at a threshold of 10, as well as a large number of misclassifications of land picture elements as water, and an 8-percent identification at 2.3 (figure 4-13d). Again, the least turbid areas of the Houston scene were identified using the statistics from the most turbid areas of Lake Livingston.

Extension of Steinhagen Lake statistics from the same day and the previous day to Lake Houston resulted in a higher degree of identification than either of the previous two lakes. Seven class statistics were used in each extension. The same problem existed as with the Livingston to Houston extension, in that one of the minor edge cell classes from Steinhagen misclassified a high number of land picture elements as water at the threshold of 10.

The same-day extension resulted in a 96-percent classification at a threshold of 10, but also a large number of classifications of land picture elements. At the 2.3 threshold, the classification resulted in a 63-percent identification (figure 4-13b). Extension from the previous day resulted in a 92-percent classification at a 10 threshold, but most of the land picture elements were also classified as water. The 2.3 threshold resulted in a 45-percent identification (figure 4-13a). The misclassification error at the 10 threshold could be entirely removed by eliminating one of the edge-cell cell classes and only extending with six classes.

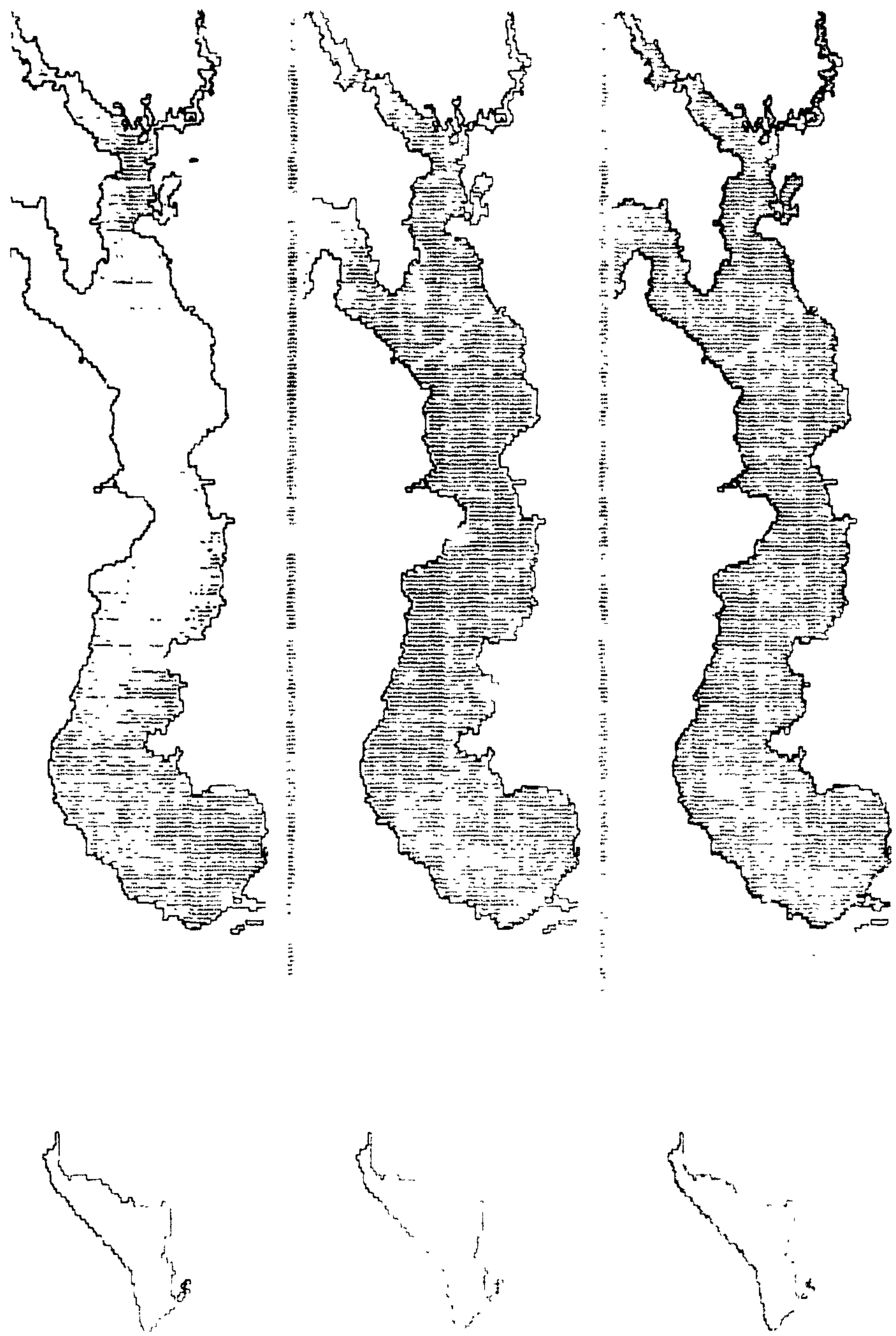
Temporal extension was attempted for Lake Houston over a period of 36 days (August 29th to October 4th). The statistics, previously reported in this section, from one, five, eight, and 14 training fields were used in this extension as both composite and separate classes.

The physical condition of Lake Houston and Sheldon Reservoir changed over this 36-day period. Rainfall increased the area of each of these lakes with no significant effect on the turbidity of Sheldon, but an increased level of turbidity on Houston which shows up as a larger "mixing bowl" area of the lake and extended turbidity inflow from the west fork.

The 36-day extension experiment followed the same format as the extension experiment within Lake Houston. The first step was to produce a gray map of the lower data values in the infrared channel to determine the location and outline of the lake (figure 4-7a). An ISOCLS map was then printed to determine the relative brightnesses over the lake on this day (figure 4-8d). The initial extension was the statistics from the single training field. The results were an identification of 29 percent at a threshold of 10 (figure 4-14a) and 5 percent at a threshold of 2.3. The areas identified were the southern end (the location of the original training field) and the main section of the East Fork.

The next extension involved the statistics from the five training fields, both as a single class and as separate classes. Both approaches led to similar results. The single class approach at a threshold of 10 resulted in

S-73-29861



(a) (b) (c)

Figure 4-14.- Signature extension results for Lake Houston data of October 4, 1972.

73-pe· ̵nt identification (figure 4-14b) and 29 percent at a threshold of 2.3. The separate class approach led to a 72-percent identification at a threshold of 10 (figure 4-10b) and 25 percent at a threshold of 2.3.

The extension of the eight training fields, both as a single class and as separate classes, resulted in the following levels of identification. The single class approach resulted in 87-percent classification at a threshold of 10 (figure 4-14c) and 55 percent at a threshold of 2.3. The separate class approach resulted in 77 percent at a threshold of 10 (figure 4-10d) and 20 percent at a threshold of 2.3. The area not identified was again the turbid West Fork. The increase in turbidity of this fork over the 36 days left no prior training field with applicable statistics.

The 14-class approach also resulted in poor identification of the West Fork of Lake Houston. The identification at a threshold of 10 was 95 percent (figure 4-8c) and at a threshold of 2.3 resulted in 41 percent. Partial results of these various extension experiments are condensed in tables 4-I, 4-II, and 4-III.

TABLE 4-I.- EXTENSION EXPERIMENT WITHIN LAKE HOUSTON

No. training fields	No. classes	Percent water identification	
		[a]T = 10	T = 2.3
1	1	[b]71	37
5	1	84	62
5	5	90	60
8	1	94	71
8	8	95	51
0	14	101	74

[a]Threshold.

[b]100% = Picture elements identified as water by the density slicing of figure 4-7a.

TABLE 4-II.- 36-DAY EXTENSION EXPERIMENT

No. training fields	No. classes	Percent water identification	
		[a]T = 10	T = 2.3
1	1	[b]29	5
5	1	73	29
5	5	72	25
8	1	87	55
8	8	77	20
0	14	95	41

[a]Threshold.

[b]100% = Picture elements identified as water by the density slicing of figure 4-7a.

TABLE 4-III.- EXTENSION EXPERIMENT FROM OTHER SITES

Site	Classes	Percent water classification	
		[a]T = 10	T = 2.3
Somerville	2	[b]6	4
Livingston	5	58	8
Steinhagen (29th)	7	96	45
Steinhagen (28th)	7	92	63

[a]Threshold.

[b]100% = Picture elements identified as water by the density slicing of figure 4-7a.

There were some interesting results from these extension experiments besides the identification statistics. The first was that no significant error was encountered (1 percent) when attempting to extend within a site and extended over time for the same site.

The features of the site were extended to the same areas as before. Logical shifts followed the expected changes in the target over the period of time. The errors in identification occurred only when borderline classes were extended from one site where they were identified as "edge-type" cells, to a different site where one of the statistics began to identify cleared areas at the larger values of the threshold. A future approach would be to input statistics for other than water sites, which might eliminate a portion of this misclassification.

Another interesting aspect was that prior knowledge of Lake Houston was required to properly place the training

fields to identify the lake. If there were no knowledge of water types, it would probably have been necessary to approach water identification through the use of a single training field in the main body of the lake. This would have resulted in a 70-percent identification of water picture elements (threshold of 10), with no identification of the West Fork or Sheldon Reservoir. In this case it would have been much better to use a "density slice" of channel 4 with gray levels of 12 and less. Gray levels of 5 or less would be acceptable if the interest were in large impoundments with little emphasis on the edge picture elements.

The highest classification accuracy was obtained through the density slice of the infrared channel, ISOCLS, and LARSYS-CLASSIFY with 12 input classes. The density slice was the easier approach to identify major water bodies. ISOCLS poses a problem in that the statistics of the various classes must be studied and an arbitrary decision made to specify which classes were water (e.g., any class with gray levels in channel 4 of 14 or less). The LARSAA-CLASSIFY worked well with artificial statistics (not developed through training fields) derived from a previous ISOCLS output.

The use of training fields required very selective choosing of training field locations, which was made easier through study of the ISOCLS output. Even with selective choosing of training fields, it was difficult, if not impossible, through the CLASSIFY routine and a meaningful threshold to identify water of differing turbidity levels from that of the training fields. This was evident in the inability to identify either Lake Livingston or Lake

Somerville with the Lake Houston training field statistics. Their turbidity levels were 25 percent of that of the least turbid areas of Lake Houston. Higher turbidity levels also caused problems for identification. Identifying the highest turbidity level of Steinhagen Lake (same day) was not possible; neither was identifying the west fork of Lake Houston on the 36-day extension. In each of these cases, the turbidity level of the area which was not classified was above that of the level of turbidity for the areas where training fields were selected. The sites separated into two groups: low turbidity (Livingston, Somerville, and Sheldon) and high turbidity (Houston and Steinhagen). Signature extension between these groups was almost impossible. The further down one proceeds on the hierarchy of target features, the more precise the statistical requirements are and also the more likely general areas of the overall feature are to be missed. In the cases studied, water in excess of 5 surface acres was extremely easy to separate from other targets in spectral space. However, once the identification was approached through the use of training fields, there was a need for being very selective in the choice of the training field to assure representation of all types. Otherwise, thresholds had to be manipulated, as well as training fields introduced for features not in the hierarchy (e.g., land features). The ability to identify the target improved as the study increased from one to eight training fields, but signature extension improved only slightly because of the nonexistence of a suitable area for training field selection in order to extend to certain sites.

Data arrived late in the study for the initial attempt at a 90-day signature extension. The results for the same experiments which were used on same-day and 36-day extensions are shown in table 4-IV. These data were only for total identification, and no attempt was made to ascertain whether the areas identified by each training field had shifted.

TABLE 4-IV.- 90-DAY EXTENSION EXPERIMENT FOR LAKE HOUSTON

No. training fields	No. classes	Percent water identification	
		[a]T = 10	T = 2.3
1	1	[b]13	0
5	1	31	1
5	5	27	1
8	1	57	2
8	8	19	0
0	14	63	1

[a]Threshold.

[b]100% = Picture elements identified as water by the density slicing of figure 4-7a.

One noticeable result has emerged. Separate classes for each training field had a higher rate of identification for same-day extension than did the single class for the combined training fields. The result was the opposite under the 36- and 90-day extensions, with the single-class approach having the higher rate of identification. This was anticipated, since the variation was probably greater

with the training fields combined than it would have been for the individual training fields.

Also as expected, identifications using all methods decreased over the 36-day extension and decreased further over the additional 54 days. This is illustrated in figure 4-15. The maximum identification under all methods using a threshold of 2.3 was only 2 percent.

Because of the late acquisition of these data, there was no attempt to color code the maps which were generated by the 90-day extension. An initial assessment indicated that the extremes of the turbidity range (West Fork and Sheldon Reservoir) were the areas consistently missed in the classification.

4.6 ISOCLS EXTENSION

An extension experiment was performed on data from two passes over Lake Livingston using the ISOCLS program. The August 29, 1972, data (scene 1037-16244) and the October 4, 1972, data (scene 1073-16244) were used. The ISOCLS program generated clusters for the August 29th frame and the clusters were then used as input for the October 4th frame. The program was allowed to iterate twice. The first iteration assigned every pixel to one of the clusters from the earlier frame, and the second iteration contained new cluster centers which were derived from the data assigned in the first iteration. The changes from the first to the second iteration were minimal, comprised primarily of slight shifts in the location of the

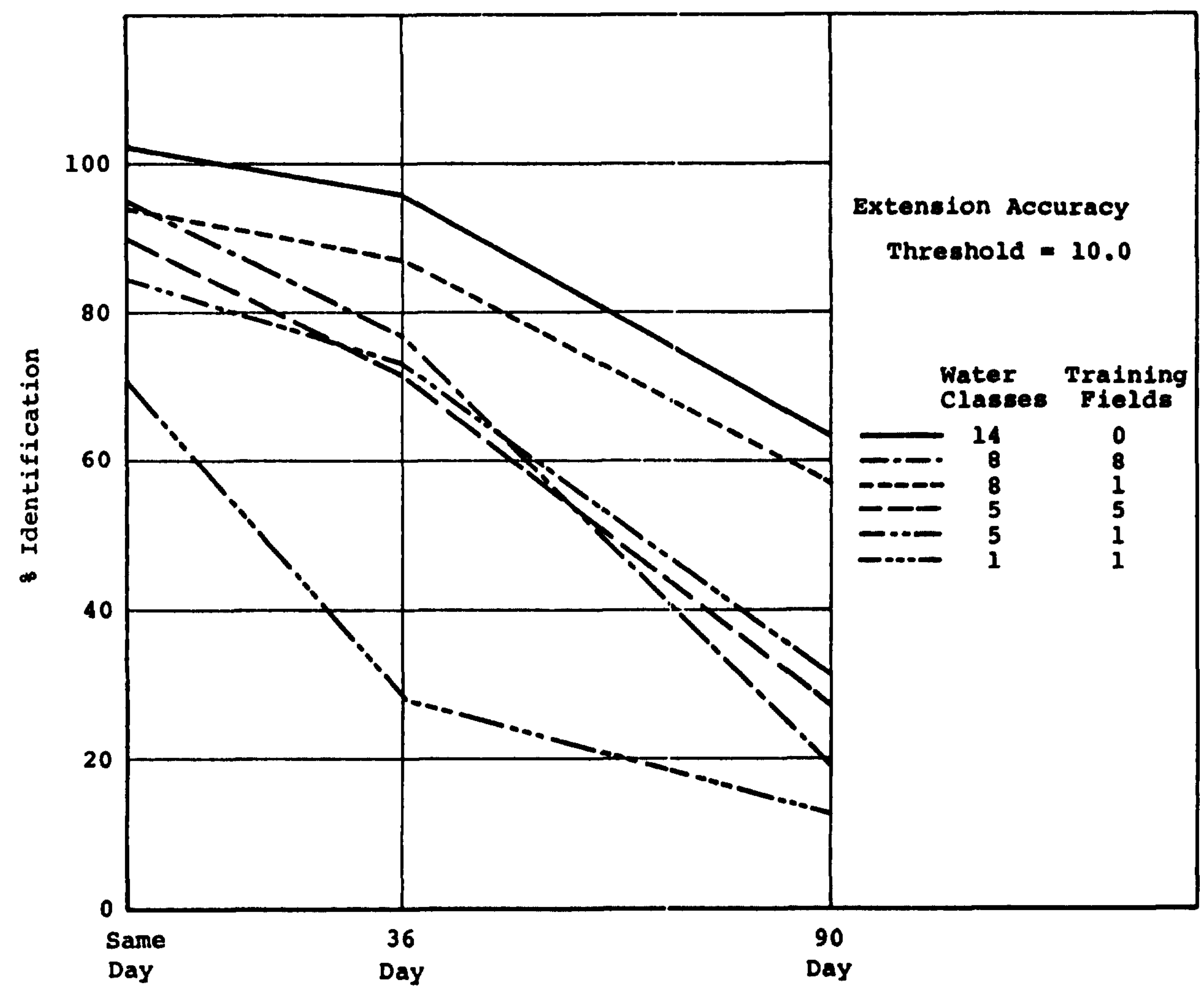
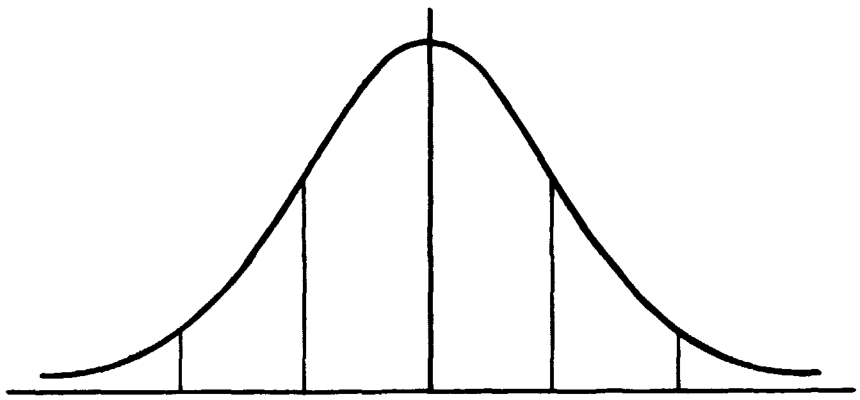


Figure 4-15.- Water identification percentages for signature extensions up to 54 days.

cluster centers. One of the original 16 clusters was deleted because only three pixels were assigned to it. Of the original 16 clusters, three represented water. The remainder were other features in the scene and were not examined in any detail. The water was so well separated from the rest of the data in spectral space that the water assignments were correct, even if the cluster centers were off by one or two data levels. The other scene features were closer to one another and tended not to have distinct boundaries in spectral space. Thus, a slight shift of the entire data set in spectral space placed pixels into adjacent clusters rather than into the correct ones. No attempt was made to investigate this type of behavior for targets other than water because of a lack of ground-truth data for the area. The ground-truth collection had been limited to the specified water targets.

4.7 THRESHOLDING EXPERIMENTS

Thresholding can best be explained by first looking at the following simple unimodal, univariate, normal distribution. Approximately 66 percent of the items taken in a sample are included in $\pm 1\sigma$ about the mean, $\pm 2\sigma$ includes approximately 95 percent, and $\pm 3\sigma$ includes about 99 percent.



From the above diagram, if all items that fall outside of ±2σ were to be threshold, all items with a value between 0 and 2 would be retained, and all other items discarded.

In the multivariate case, such as a LARSYS-type classifier applied to four channels of data, the problem becomes more difficult to understand, but the principle remains the same. Basically, the threshold value determines how close the four-channel data values of a pixel have to be to the respective means of the four channels, as determined by the training field data, before the pixel is classified as being the same type of item as the training fields.

Empirically-derived values of threshold versus the percent of classified pixels within a training field (for agricultural products) were used as first approximations for classification of water.[1] In general, these were

[1] "Empirical Distribution of Quadratic Form Used for Thresholding," by W. G. Eppler, LEC/HASD No. FSD-001, November 1972. LEC Job Order 81-173.

found to be accurate enough to be used for the purposes of this study. They are

Threshold	Percent
2.3	66
3.0	80
4.7	95
6.5	99
10.0	100

Figure 4-16 shows the results of four of these five thresholds as applied to Lake Houston. The rectangle at the bottom of the lake defines the training field used. Actual classification statistics are

For a threshold of -	Percent of training field classified	Percent of lake classified
2.3	68	37
3.0	81	45
4.7	93	57
6.5	99	65

Figures 4-17, 4-18, and 4-19 show how the threshold affects classification using varying numbers and types of training fields, and classifying Lake Houston into separate and combined classes of water.

4.8 WATER DETECTION

With water bodies as the primary target, attention naturally turned to detection of water in the ERTS scenes using computer-compatible tape.

(a) Threshold=2.3 (b) Threshold=3.0 (c) Threshold=4.7 (d) Threshold=10.0

Figure 4-16.- Effect of different thresholds using one training field and one class to identify water. (Lake Houston data collected August 29, 1972.)

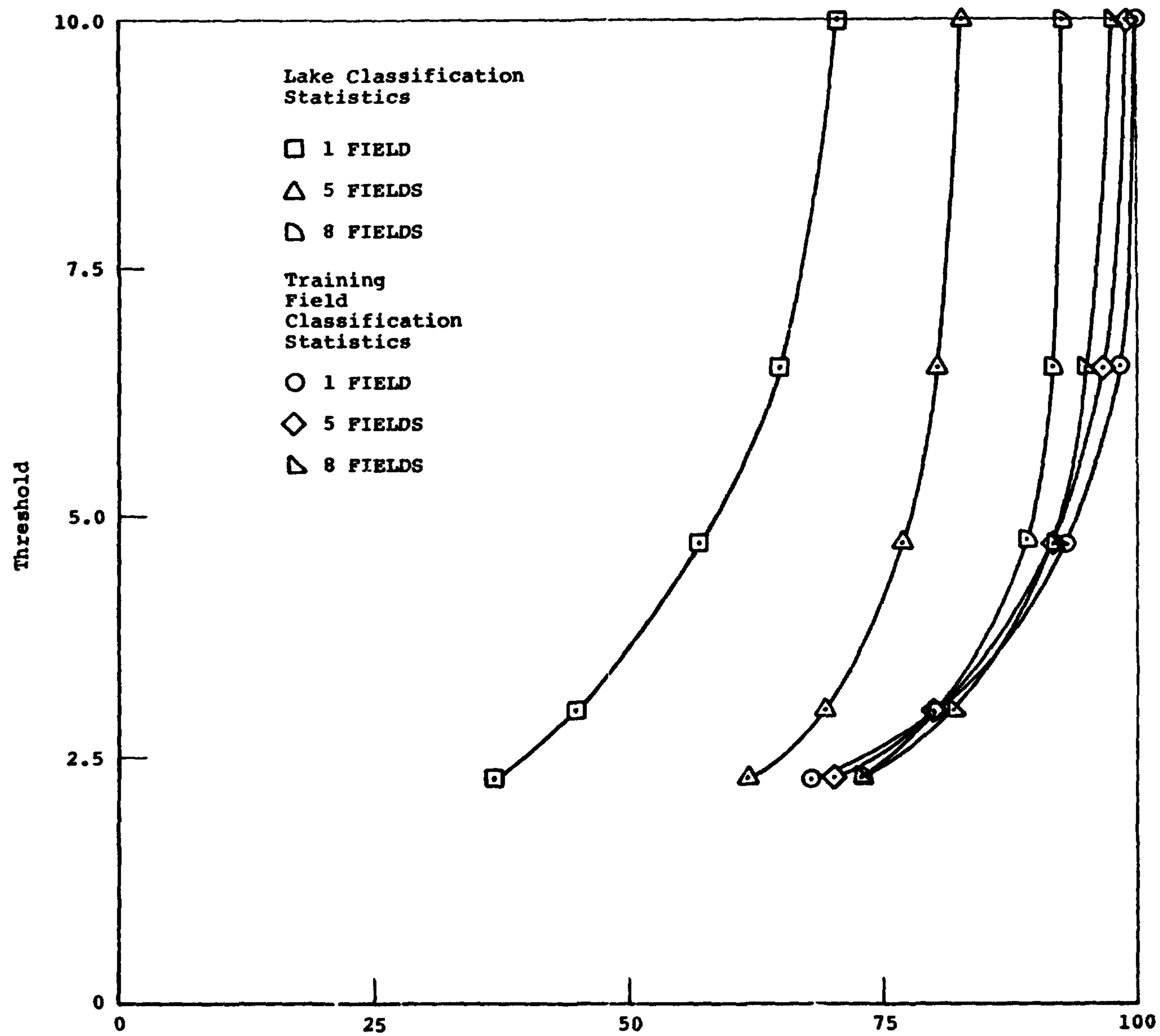


Figure 4-17.- LARSYS classification of Lake Houston with various training fields assuming one water class.

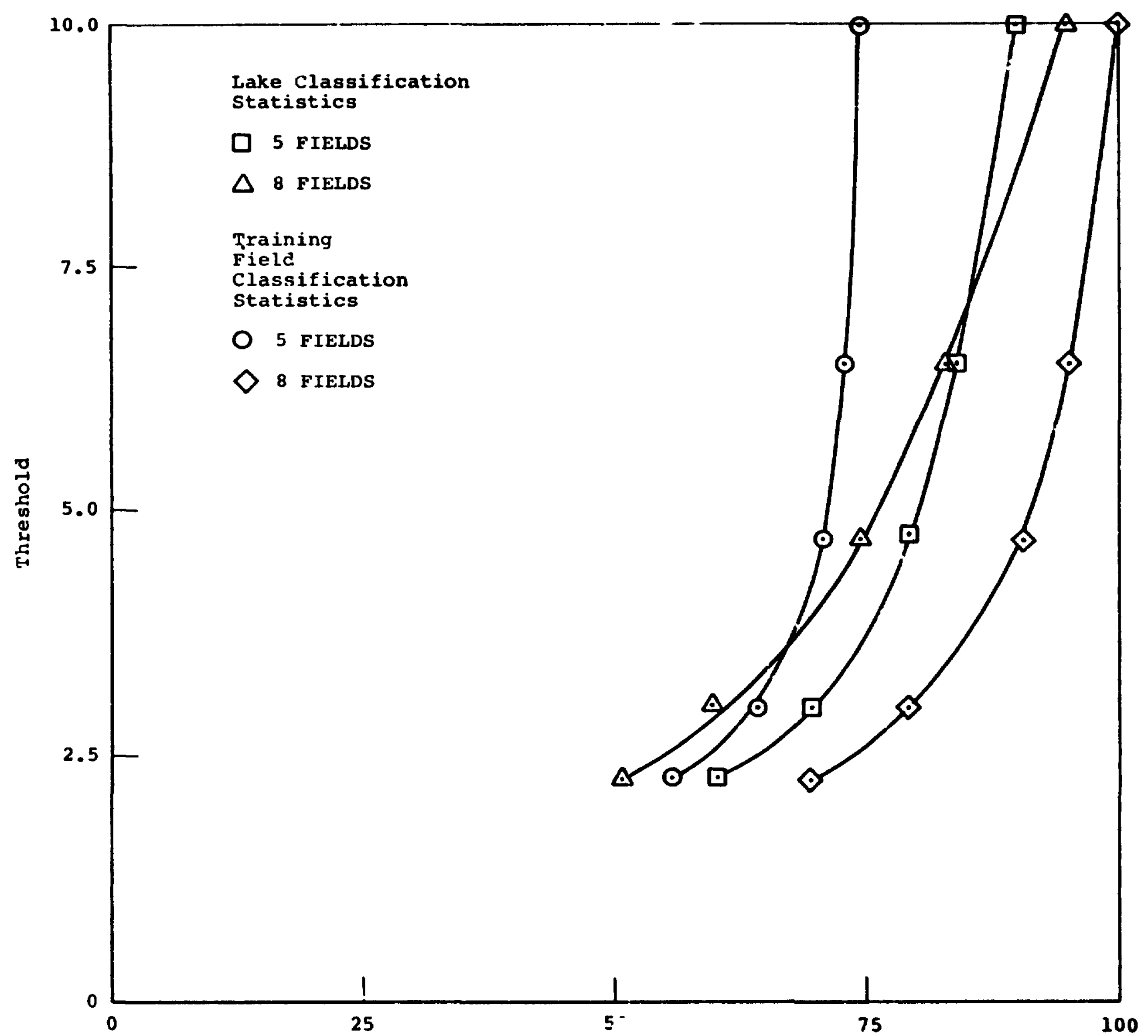


Figure 4-18.- LARSYS classification of Lake Houston with five and eight training fields, assuming five and eight classes of water.

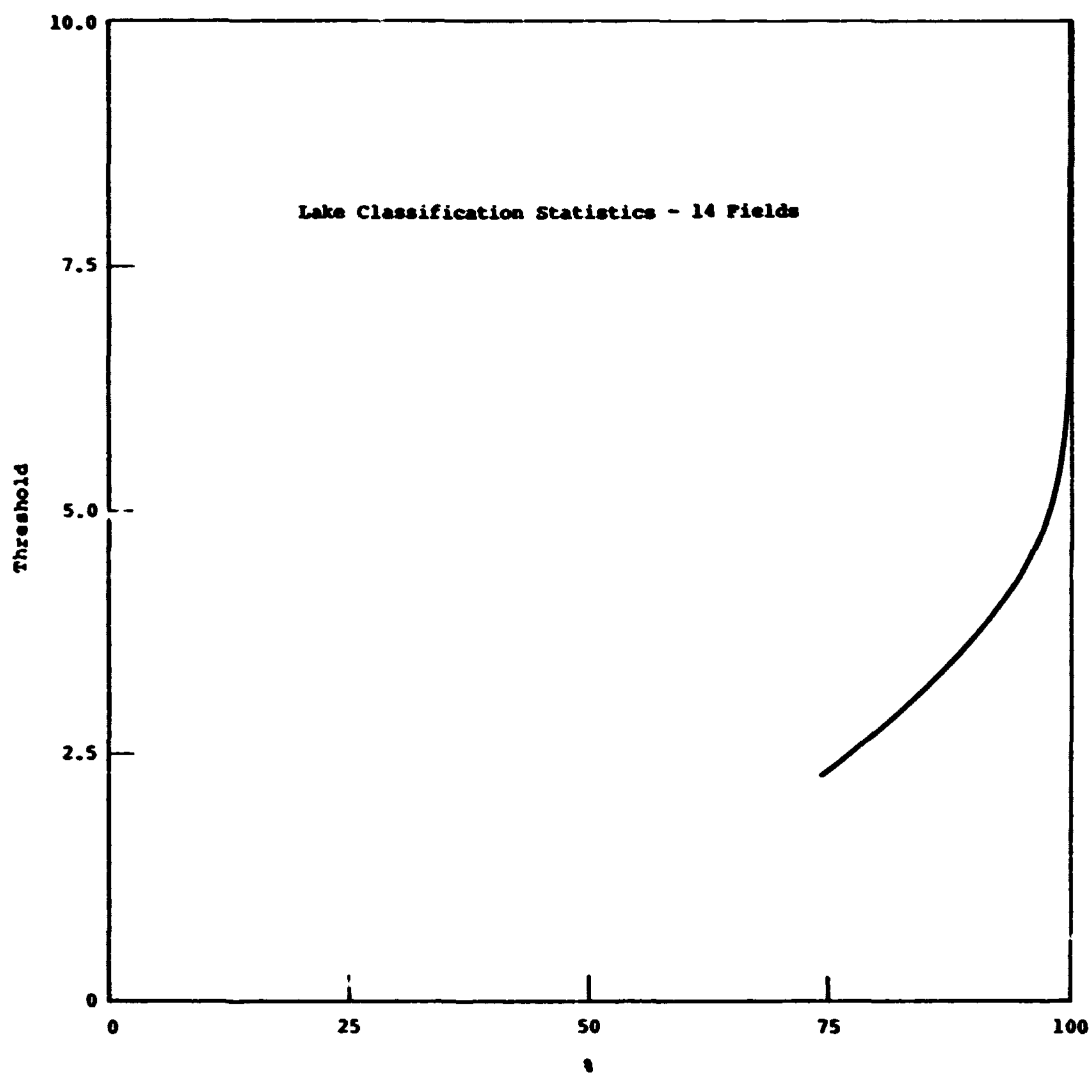


Figure 4-19.- LARSYS classification of Lake Houston with training field data obtained from ISOCLS for 14 types of water.

Quite early in the examination of the ERTS-1 data, low values in the infrared channel 4 band were noticed to be associated with water. Both the Monterey Bay and the Lake Somerville data of July 25, 1972 indicated that low values in channel 4 indicated water. All but a few data points in Lake Somerville were in the 0 to 4 range for channel 4.

Increasing the maximum data values from 4 to about 12 in channel 4 filled in a few pixels around the edges of Lakes Somerville, Livingston, and Houston, and a few isolated groups of low data values occurred away from the large lakes. An examination of aerial photography disclosed that the isolated groups were ponds of water of a few acres. For the scenes examined (August 29 and 30, 1972; October 4, 1972), a pixel with a data value of 12 or less in channel 4 had water in the field of view.

Attempts to use the 0 to 12 or even 0 to 9 criterion on the October 23, 1972, data resulted in large areas of lowlands being identified as water. These areas were water, but only a few inches deep, with a great deal of vegetation protruding above the water's surface. To eliminate the wet fields from the water identification would require the allowable data values to be restricted to the 0 to 5 or 0 to 6 range. Such a restriction sacrifices many of the edge pixels around the large lakes and ponds, but the main body of water is still detected.

Turbid water was noticed to have higher data values in channel 1 and slightly higher data values in channel 4 than clear clean water. Consec ly, the channel 4 data values

could be allowed to go higher than 5 or 6 if the channel 1 value was high. Plots were made of channel 1 data versus channel 4 data to determine if a simple curve could be placed between the water and the nonwater data points. The first few tests were of straight lines which passed through the origin and had slopes in the vicinity of 4 (channel 1 data value divided by channel 4 data value).

When the slope was less than 4, the small turbid ponds were detected, but there were false alarms in the wet lowlands. When the slope was more than 4, the false alarms were eliminated, but the small, turbid ponds were also lost. The solution to that problem was to move the straight line away from the origin so that it would have a slope of less than 4, but would still separate the deep water from the wetlands at a data value of 5 or 6 in channel 4. An intercept of 8.5 (when the channel 4 data value was 0) and a slope of about 2.8 was tried, and this value retained the muddy ponds while eliminating the false alarms.

Because there were so few data points for water (only about four-tenths of 1 percent, even when a large lake such as Lake Somerville was present), it was not practical to try to refine the location of the straight line. Also, it was not possible to determine what nonlinearity might do to improve the detection of water.

5.0 CONCLUSIONS

1. The spectral signature of water was very stable and was well separated from all other elements of the scene in spectral space.

2. The signature derived from one body of water will only extend to another body of water which has the same turbidity.

3. A class called "water", which includes water of all possible turbidities, occupies a region of spectral space which is incompatible with the methods of describing classes in both LARSAA and ISOCLS.

4. Most of the information necessary for separating water from nonwater is in the channel 4 data.

5. There are two major sources of signature variability, differences in the target itself and differences in the illumination level caused by different solar elevation angles.

6. Changes in the atmosphere and residual miscalibration of the data are minor sources of signature variability.

Lyndon B. Johnson Space Center
National Aeronautics and Space Administration
Houston, Texas, April 24, 1974
641-14-07-50-72